Evaluation of pyruvate kinase as a potential phylogenetic marker: Studies on the pyruvate kinase of the archaebacterium *Thermoplasma acidophilum.*

Simon Potter, B.Sc. (Hons)

Thesis presented for the degree of Doctor of Philosophy
University of Edinburgh
1993

"Alas, how terrible is wisdom when it brings no profit to the man that is wise!"

Sophocles, *Oedipus the King*

"It was the best of times, it was the worst of times."

Dickens, *A Tale of Two Cities*

Declaration

I hereby declare that this thesis has been composed by me, that it has not been accepted in any previous application for a degree, and that the work of which it is a record has been carried out by me.

SIMON POTTER B.Sc. (Hons)

Acknowledgements

My sincere thanks are due to:

My supervisors Dr. Linda A. Fothergill-Gilmore and Prof. Richard P. Ambler for their enthusiasm, patience and guidance throughout my time in Edinburgh.

Messrs. Malcolm McGregor and Derek Shaw for their technical help with the archaebacteria reactor.

Dr. David Hulmes for his help with the electron microscopy.

Andy Cronshaw, Linda Kerr and Steven Peacock for their tireless sequencing and amino acid analysis assistance.

Drs. Malcolm White, Fiona Stuart, Lluis Ribas, Jacqueline Nairn, Caroline Colville, Craig Simpson, Gary Patterson, Steve Johnstone, Andrew Ryle and Graeme Pettigrew for numerous (and occasionally helpful) discussions.

The many people who have found themselves in lab. 327 over the years for many a chuckle in dark times.

Others outside of the walls of 327 too numerous to mention.

Prof. Patrick P. Dennis for offering me a post-doc. (see you in April, honest!).

My family, for their love, support and hard cash.

And finally, to Carol for more than I can fit on this page (or in this thesis).

Thanks a 10^6.

Abbreviations

A	-	Adenine
ADP	-	Adenosine-5'-diphosphate
AMP	-	Adenosine-5'-monophosphate
ATP	-	Adenosine-5'-triphosphate
bp	-	Base pairs
C	-	Cytosine
C-	-	Carboxy-
CTP	-	Cytosine-5'-triphosphate
d-	-	Deoxy-
DMSO	-	Dimethyl sulphoxide
DNA	-	Deoxyribonucleic acid
dNTP	-	Deoxynucleoside-5'-triphosphate
ds	-	Double stranded
DTT	-	Dithiothreitol
EC	-	Enzyme commission number
EDTA	-	Ethylenediaminetetraacetic acid
G	-	Guanine
g	-	gramme
GTP	-	Guanine-5'-triphosphate
HPLC	-	High performance liquid chromatography
k	-	kilo
kb	-	kilobases
K_m	-	Michaelis constant
l	-	Litre
mA	-	Milliampere
mg	-	Milligramme
ml	-	Millilitre
Mr	-	Relative molecular mass
µg	-	Microgramme
µl	-	Microlitre
N	-	Amino-
NADH	-	Nicotinamide adenine dinucleotide (reduced)
OD	-	Optical density
ORF	-	Open reading frame
PAGE	-	Polyacrylamide gel electrophoresis
PEP	-	Phosphoenolpyruvate
rpm	-	Revolutions per minute

SDS	-	Sodium dodecyl sulphate
SDW	-	Sterile distilled water
ss	-	Single stranded
T	-	Thymine
Tm	-	Melting temperature
Tris	-	Tris(hydroxymethyl)aminomethane
TTP	-	Thymidine-5'-triphosphate
UV	-	Ultraviolet
V	-	Volts

Amino acids

A	Ala	Alanine
B	Asx	Asp/Asn
C	Cys	Cysteine
D	Asp	Aspartate
E	Glu	Glutamate
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
Z	Glx	Glu/Gln

Figures

Tables

CONTENTS

Abstract

Molecular evolution is a relatively new field of science that brings to evolutionary biology the powerful methods of both molecular biology and computational science. The field is based upon comparative data obtained from molecular studies of DNA, RNA and proteins isolated from a wide variety of organisms, and has led to the controversial reclassification of living organisms into the three "domains" - eucaryotes, eubacteria and archaebacteria (eucarya, bacteria and archaea). The molecules chosen for these studies are known as phylogenetic markers and should fulfil the following criteria: ubiquity, ease of isolation and constancy of function. In order to extend the field of research and to increase the accuracy of interpretation of the data, new phylogenetic markers are constantly required.

In this study the glycolytic enzyme pyruvate kinase was chosen as a potential phylogenetic marker. It was considered to be particularly appropriate for two reasons; first, it fulfils the criteria outlined above and secondly, there is a wealth of structural information for the enzyme from various sources available in the sequence databases. The pyruvate kinase chosen for study was that of the thermoacidophilic archaebacterium, *Thermoplasma acidophilum*, because whilst there is a great deal of information in the databases on pyruvate kinases isolated from organisms within the eucaryal and bacterial domains, no sequence information previously existed for an archaebacterial equivalent. The aim of this study, therefore, was first to characterise the enzyme to ensure that the constancy of function requirement was fulfilled, and secondly to obtain sufficient primary structure information to facilitate a realistic evaluation of the potential of pyruvate kinase as a phylogenetic marker.

The pyruvate kinase was purified to homogeneity using a series of chromatographic steps, and then characterised with respect to its physical and kinetic properties. The enzyme has a native Mr of 250 K, and a subunit Mr of 60 K. It exhibits typical K_m values towards its substrates PEP and ADP, and is allosterically regulated by AMP. It is one of the most thermostable pyruvate kinases yet isolated, being active at up to 90^{o}C. Initial sequencing attempts were frustrated by the chemical blockage of the N-terminus of the enzyme, and hence it was cleaved both chemically and proteolytically into peptide fragments, which were then sequenced by automated Edman degradation. The sequences of these internal peptides were then used in conjunction with a codon usage table derived from the citrate synthase gene of *T. acidophilum* to design a number of oligonucleotide probes. These probes were then a) fluorescently labelled and hybridised to Southern blots of restriction digests of *T. acidophilum* genomic DNA, and b) used

as primers for the polymerase chain reaction.

Labelled oligonucleotide probes O1 and O2 identified the same 2 kb EcoR1/HindIII fragment of genomic DNA, whilst probes O1 and O4 (which correspond to conserved regions of pyruvate kinase sequence) generated a 0.6 kb PCR product. Both of these stretches of DNA were cloned into plasmid vectors for dideoxy sequencing. The nucleotide sequences obtained were translated into their corresponding polypeptide sequences, which were subjected to multiple alignment and secondary structure prediction analyses. From these studies, it appears that the archaebacterial enzyme sequence differs widely from that of its mesophilic counterparts whilst retaining some regions of structural and/or catalytic importance.

Chapter 1: Introduction

CHAPTER 1: INTRODUCTION

1.1 Molecular evolution

1.1.1 Historical perspectives: the use of molecular chronometers and the concept of the archaebacteria

Prior to the 1970's, taxonomic studies of organisms and their phylogenetic interrelationships were based exclusively upon complex morphologies and large numbers of phenotypic characters. The evolutionary relationships between organisms were then extrapolated from these taxonomic studies using evidence garnered from the fossil record as a guide. Studies of this type led to the acceptance of the five-kingdom phylogeny proposed by Whittaker (1959), which was subsequently developed and enlarged upon by other groups (for a review, see Margulis and Guerrero, 1991). In this proposal, organisms were split into five equally ranked groups or 'kingdoms': metabionta (animals), metaphyta (plants), fungi, protocista and monera (the bacteria). A major problem with analyses of this kind, however, is that microbial morphologies are too simple and/or uninterpretable to provide a valid taxonomy and, therefore, evolutionary considerations were necessarily confined to complex organisms. The histories of such higher organisms at best cover only the most recent 40% of evolutionary time, and therefore little evolutionary information could be deduced for the previous 60% (a period of time amounting to nearly 2 billion years).

In the 1970's, however, the sequencing revolution led to the established dogma being challenged. The basis for the challenge was a series of comparative studies of the protein and DNA sequences of conserved macromolecules isolated from various prokaryotic and eukaryotic sources. These molecules became known as *phylogenetic markers* or *molecular chronometers* and were required to fulfil the following criteria: ubiquity, ease of isolation and constancy of function (which would ensure that any mutations would be selectively neutral). In a pioneering study of this series, Carl Woese of the University of Illinois chose the 16S/18S ribosomal RNA molecule as a phylogenetic marker because, in addition to satisfying the requirements of such a

molecule, the rRNA is thought to possess regions which evolve at different rates, thereby giving the molecular chronometer both a 'minute hand' (the faster mutating regions) for detecting close taxonomic relationships, e.g. within phyla, and an 'hour hand' (the slower mutating regions) for detecting relationships between organisms which diverged from one another much earlier.

Phylogenetic trees are diagrams which illustrate the taxonomic relationships between organisms where any two organisms are separated by a line, the length of which is determined by the perceived difference between them. To produce such trees using the 16S/18S rRNAs as marker molecules, Woese originally used a process known as oligonucleotide cataloguing (Woese and Fox, 1977). 16S/18S rRNA molecules were digested with ribonuclease T1 and the resultant oligonucleotides were separated electrophoretically and sequenced. This generated a catalogue of information which could then be used to determine the number of nucleotide differences between the rRNA sequences, and from that information trees could be constructed.

Trees of this type showed up the expected groupings of the eubacteria, or true bacteria, and the eukaryotes, which could be further divided into the four kingdoms of higher organisms - metabionta, metaphyta, protocista and fungi (Whittaker, 1959). As the studies continued, however, adding more and more rRNA sequences isolated from further organisms, it was noted that, in addition to the above groupings, there was a third, large cluster developing. All the species in this cluster were prokaryotic in nature and yet their rRNA sequences, whilst being similar to one another, were as different to those of other prokaryotes as they were to those of eukaryotic species. In 1977, Woese proposed that this new, third grouping of organisms should be treated as a distinct taxonomic kingdom and christened them the *archaebacteria*. A more recently derived phylogenetic tree illustrating the three major groupings of organisms is shown in Fig. 1.1.

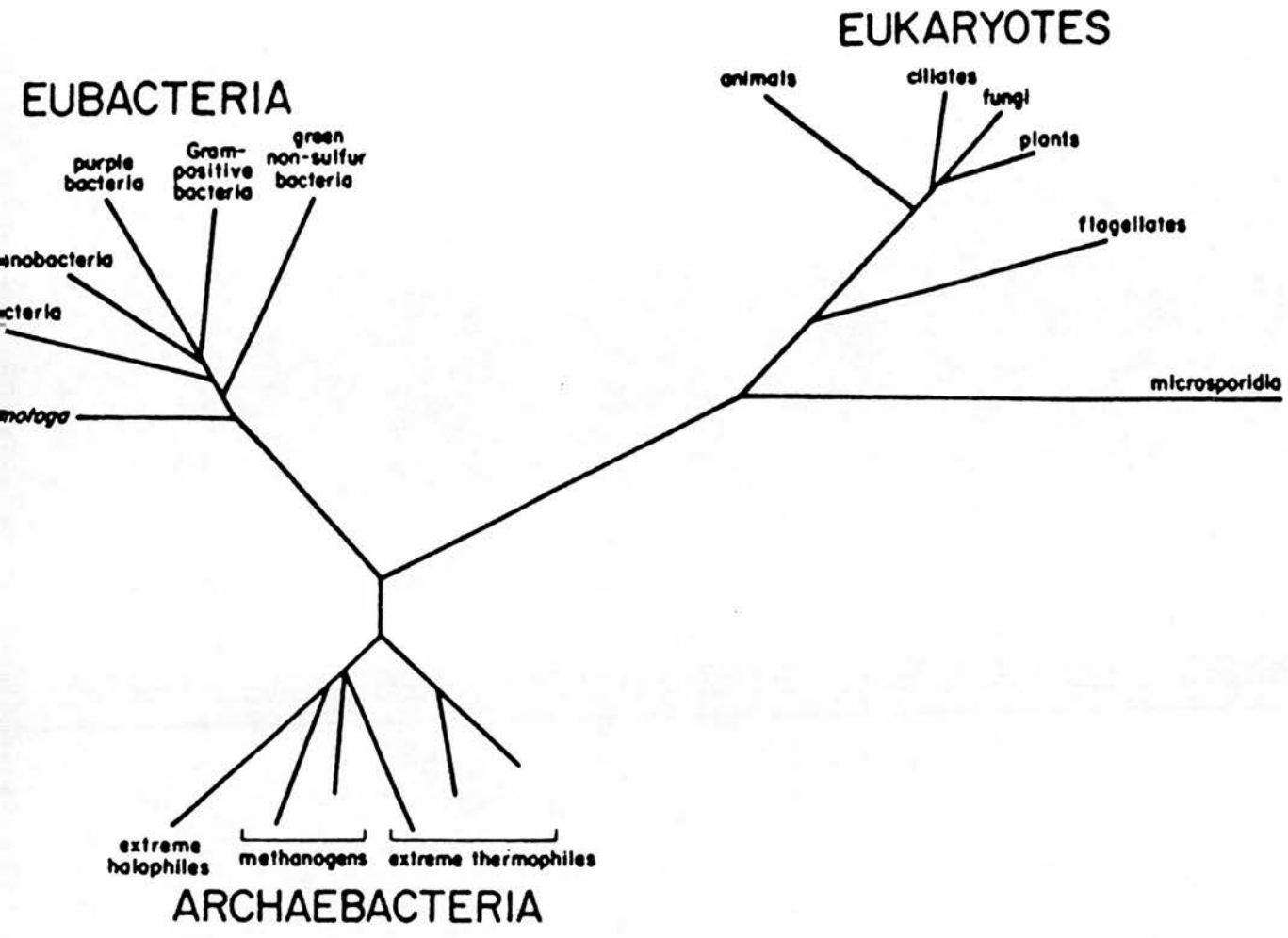


Figure 1.1: Universal phylogenetic tree (the 'archaea' tree), showing the three domains of organisms defined according to Woese *et al.* (1990). Branching order and branch lengths are based upon rRNA sequence comparisons.

1.1.2 Use of computers and basic concepts

The approach of oligonucleotide cataloguing has since been superseded by the use of complete sequence studies. Since the recent merging of the main nucleotide sequence databases (EMBL and GENBANK), the amount of easily available sequence information has risen to over 100 million bases of DNA. Many complete protein and RNA sequences are available within this source and can be accessed and aligned together using several computer algorithms. One such program is CLUSTAL (Higgins and Sharp, 1986) and it is this program which has been employed to generate the multiple sequence alignments included within this thesis.

CLUSTAL is a multiple alignment package which produces maximum identity between sequences by introducing small gaps into them. Obviously gapping cannot be unrestrained and, therefore, the program incorporates a penalty system. This ensures that the introduction of a gap is ten fold less likely than the acceptance of a single amino acid change by the program. Penalties can also be introduced manually to ensure that insertions and deletions do not fall within elements of secondary structure such as α-helices or β-sheets. Such secondary structural information is fed into the program based on the known three-dimensional structure of a representative member of the family of proteins which are being aligned.

If gapping were not allowed, two random sequences would be about 5-6% identical. Since CLUSTAL attempts to maximise sequence identity by the use of gaps, it is found that two entirely unrelated sequences show 20-25% identity with one another when aligned using this program. For any reliable indication of sequence similarity between the proteins aligned by this package, therefore, the percentage identity value obtained must exceed 25% - the 'threshold' value. If two proteins show percentage identities around or below the threshold value then no positive indication of relatedness can be deduced and other evidence must be obtained.

Once an alignment has been constructed, a matrix of percentage

differences (or similarities) can be derived. Evolutionary relatedness between amino acid sequences can then be expressed as 'accepted point mutations per 100 residues' or PAMs (the smaller the PAM value, the higher the degree of relatedness). This is a parameter, which was originally calculated by Dayhoff (1978), which is related to the matrix of differences and statistically takes into account the back mutations and multiple hits which may have occurred during evolution. The PAM values can then be used to construct phylogenetic trees relating the sequences to one another graphically. CLUSTAL is also capable of this function - during the course of the alignment, the program picks the two most closely related sequences as a base upon which to build up a tree.

1.2 The archaebacteria

1.2.1 The taxonomic status of the archaebacteria

Phenotypically, the archaebacteria fall into three distinct classes: the **methanogens** (e.g. *Methanobacterium thermoautotrophicum*), obligate anaerobes which reduce carbon dioxide to methane and exist in swamps and ruminant gut regions, the **thermophiles** (e.g. *Sulfolobus solfataricus*) some of which can survive at temperatures up to 110°C in boiling mud pools, hydrothermal vents etc. (see plates 1.1 and 1.2), and the **halophiles** (e.g. *Halobacterium halobium*) which have an absolute requirement for high salt concentrations (some even flourish in saturated salt (5.2 M), whilst others have a requirement for high pH, e.g. 9-10).

The archaebacteria pose several taxonomic problems to the evolutionary biologist. First, not all workers in the field believe that they deserve their separate classification outside the existing domains of eukaryotes and prokaryotes. Mayr (1991), for instance, insists that the most important delineation of living organisms must remain that which separates eukaryotes from prokaryotes - a division first noted by Chatton (1938). Since the archaebacteria are unquestionably prokaryotic in organisation, they would fall into the second of the two major groups which Mayr would recognise. Secondly, amongst even those who do

Plates 1.1 and 1.2: Archaebacterial mats at the village of Hveragerdi (near Reykjavik) in Iceland. The temperature of the water pools in the area was measured to be 80^{o}C.

share the belief that the archaebacteria do constitute a separate taxonomic grouping, there is considerable controversy as to how the different groups of organisms within the archaebacteria relate to one another.

In their most recent discourse on the taxonomy of the archaebacteria, Woese *et al.* (1990) divided the archaebacteria into two major phylogenetic groupings: [1] the crenarchaeota, comprising all of the sulphur-dependent extreme thermophiles and [2] the euryarchaeota comprising all the methanogens, halophiles and some thermophilic species (e.g. *Thermoplasma acidophilum* and *Thermococcus celer*). They went on to propose a new name for the archaebacteria, the *archaea*, giving them equal status to the other superkingdoms, or 'domains', which would be called the *bacteria* and the *eucarya* (see Fig. 1.1) Other workers disagree with these divisions. Most notable is James Lake who proposed (Lake, 1991) that the archaebacterial tree of Woese *et al.* was incorrect and that the archaebacteria do not comprise a monophyletic group (i.e. one containing a common ancestor and all of its descendants) but instead fall into three groups - the photocytes (mostly halophiles which classify with the bacteria), the eocytes (equivalent to Woese's 'crenarchaeota' which classify with eukaryotes) and the methanogens (which Lake believes to be the only 'true' archaebacteria). This classification led to Lake's development of the 'eocyte tree' (Figure 1.2). The source of the division is simply the choice of a different computer algorithm by each research group to produce their phylogenetic trees. Further, Lake uses a different part of the structure of the rRNA molecule to obtain sequence data.

Sceptics of these phylogenetic trees abound, particularly amongst traditional cladisticians such as Lynn Margulis and Ernst Mayr. Both of these workers have recently published critiques of the new tenets of molecular evolution (Margulis and Guerrero, 1991; Mayr, 1991), the essence of which is that Woese, Lake and their colleagues are placing too much emphasis on too little information. The argument is that the new phylogenies are based on too few characters of pre-weighted value, i.e. structural regions of rRNA molecules, and that single-gene phylogenies of this type invariably turn up evolutionary anomalies (a position

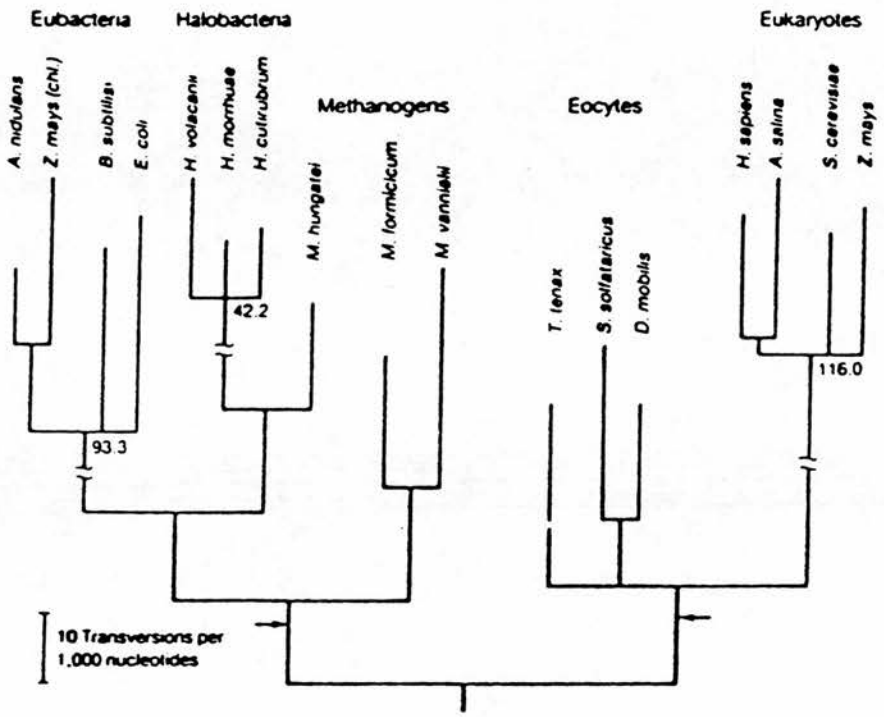


Figure 1.2: Universal rooted phylogenetic tree (the 'eocyte' tree) showing the five groups of organisms as defined by Lake (1988). Branching order and branch lengths are based on rRNA sequence comparisons.

supported by Richard Ambler, 1985). In support of their case, they cite the long held belief that the cyanobacteria were eukaryotes - the 'blue-green algae'.

Other cladisticians wonder if the choice of marker molecules used in these studies may be mistaken. Sneath (1989) suggests that genes subjected to strong selective pressures, such as those coding for rRNA molecules, would give a misleading picture of evolutionary processes and rates - they would logically seem to be constrained to evolve peculiarly slowly. He therefore proposes that pseudogenes, which do not experience such pressures because they do not produce functional gene products, might be better subjects for evolutionary studies as they could be expected to accumulate changes at a steady rate.

Zillig (1989) proposed that evolutionarily conserved proteins, rather than rRNA molecules, should be used as phylogenetic markers and produced a tree remarkably different from any seen before (Fig. 1.3). From comparative studies using genes coding for the DNA-dependent RNA polymerase enzymes from three archaebacteria (*Sulfolobus acidocaldarius*, *Methanobacterium autotrophicum* and *Halobacterium halobium*), the three eukaryotic enzymes pol 1, 2 and 3, and polymerases from *E. coli* and a chloroplast, Zillig drew two conclusions. First, that the archaebacteria are a monophyletic group and secondly, that the eucyte (or first eukaryote) "arose from the fusion of partners of which one was close to the root of the archaebacteria, the other to early eubacteria". This 'fusion' hypothesis bears some similarity to the endosymbiotic theory proposed to explain the origin of chloroplasts and mitochondria and may help to shed some light, not only on the taxonomic interrelationships of the three major groupings but also on their evolutionary relatedness. For a brief review of current thinking in archaebacterial taxonomy see Appendix 2 (Potter, 1992a).

1.2.2 The unique nature of the archaebacteria

The status of the archaebacteria as a third primary kingdom, whilst having been questioned on phylogenetic grounds, has been supported by

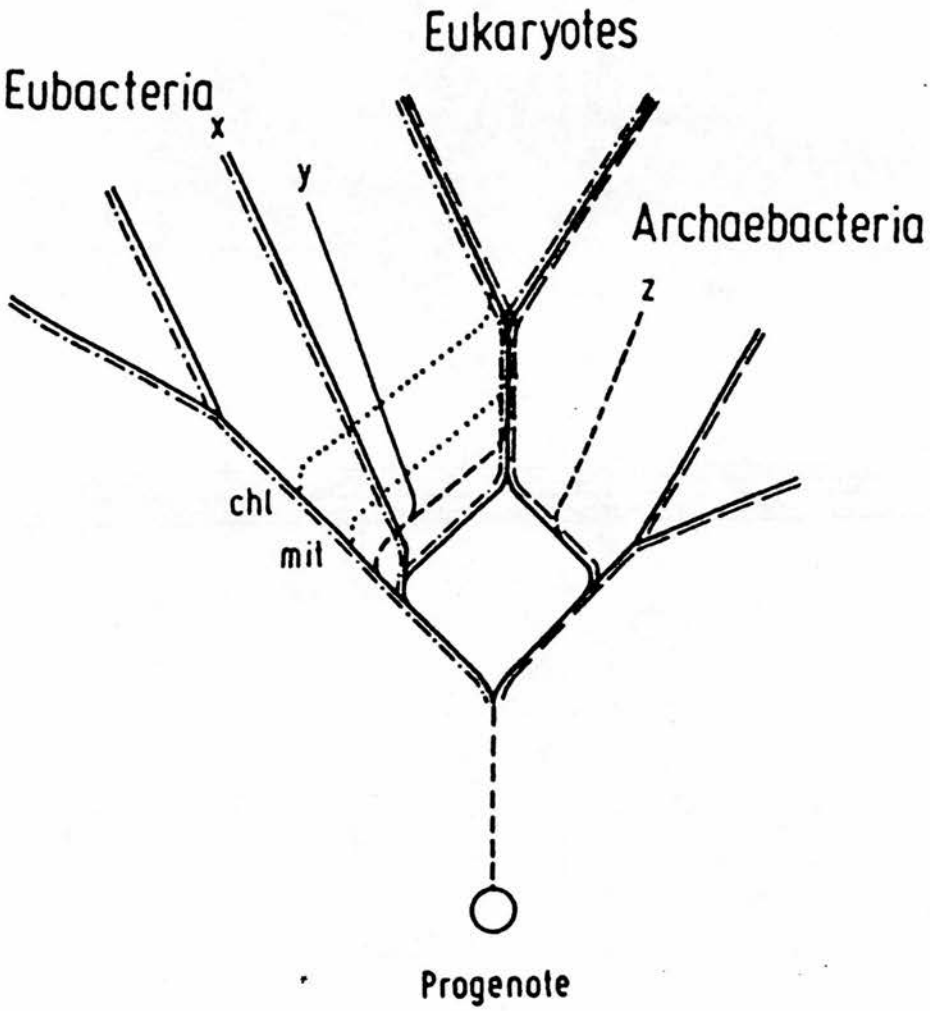


Figure 1.3: Schematic representation of the generation of the eucyte by a fusion event involving an ancestral archaebacterium and an ancestral eubacterium. The lengths of the branches are arbitrary but the branching order is based on sequence comparisons of DNA-dependent RNA polymerase subunits. Reproduced from Zillig *et al.* (1989).

many studies of their physiology and biochemistry. Table 1.1 shows the characteristics of some archaebacterial genera (Jones *et al.* 1987, Hough and Danson, 1989) whilst Table 1.2 shows unique and shared features of archaebacteria, eubacteria and eukaryotes (Woese, 1981). It can be seen from Table 1.2 that the archaebacteria are organised in a prokaryotic manner like the eubacteria whilst possessing many features which could previously have been described as being eukaryotic. Their most striking unique feature is the composition and nature of their plasma membranes. Archaebacterial cell envelopes lack muramic acid, diaminopimelic acid, D-glutamic acid, D-alanine and murein and have, in their place, N-acetylglucosamine/galactosamine and some L-amino acids. Further, the predominant glycerolipids in eubacteria and eukaryotes are glycerides (acylglycerols), whereas in archaebacteria they are glycerol ethers. The presence in archaebacterial membranes of C40-tetraethers allows the formation of rigid monolayers instead of the more usual bilayers - this may be an adaptation to life in hostile biotopes (biological niches) as such membranes are more resistant to high temperatures and, unlike esters, ethers are resistant to acid/base hydrolysis. Some typical archaebacterial lipids are shown in Figure 1.4.

For a review of the biochemistry of the archaebacteria see Danson (1988) and Potter (1992b, Appendix 2).

1.2.3 Biotechnological potential of the archaebacteria

Halophiles: Extremely halophilic archaebacteria were an early focus for interest owing to their role in the deterioration of salted food. When salted fish is stored in warm, humid conditions, a red patina (the 'pink') often develops due to the growth of some *Halobacterium* or *Halococcus* species (Shewan, 1971). Since both genera are proteolytic, considerable degradation of the material can occur. Whilst this phenomenon has been largely eliminated by the advent of refrigeration, in some cases it can be desirable. For example, Thai *Nam Pla* ('fish sauce'), a widely used condiment in southeastern Asia, is the product of the action of halobacteria on fish proteins. It is interesting to note that this may be the only example of archaebacteria consumed in the human diet.

PHENOTYPE	GENERA	RELATION TO O_2	OPTIMAL NaCl (M)	G + C (mol %)	OPTIMUM TEMP (^{O}C)	OPTIMUM pH
Halophilic	*Halobacterium*	aerobic	2.0-4.5	60-68	35-50	6.5-7.2
	Halococcus	aerobic	3.5-4.5	65	30-37	7.2
	Natronobacterium	aerobic	3.0-3.5	63-65	37-45	7.2
	Natronococcus	aerobic	3.5-4.0	64	35-40	9.5
Thermo-philic	*Thermoplasma*	facultatively aerobic	-	38-46	59-60	1.0-2.0
	Sulfolobus	aerobic	-	30-40	70-90	1.5-5.0
	Archaeoglobus	anaerobic	-	46	83	5.5-7.5
	Thermococcus	anaerobic	-	51	92	5.8
	Thermoproteus	anaerobic	-	55	88	5.5
	Thermophilum	anaerobic	-	57	88	5.5
	Thermodiscus	anaerobic	-	53	87	5.5
	Desulfurococcus	anaerobic	-	51	85	6.0
	Pyrodictium	anaerobic	-	50-62	105	5.0-7.0
	Pyrococcus	anaerobic	-	ND	70-103	5.0-9.0
Methano-genic	*Methanococcus*	anaerobic	-	29-34	30-85	6.0-9.0
	Methanothrix	anaerobic	-	52	37-60	7.4-7.8
	Methanosarcina	anaerobic	-	39-42	35-50	6.0-7.0
	Methanomicrobium	anaerobic	-	45-49	40	6.1-7.0
	Methanogenium	anaerobic	-	47-61	20-60	6.2-7.5
	Methanobacterium	anaerobic	-	32-61	37-70	7.0-8.5
	Halomethanococcus	anaerobic	2.0-3.0	49	35	7.5
	Methanohalophilus	anaerobic	0.2-2.0	ND	30-45	9.0

Table 1.1: Characteristics of some archaebacterial genera. ND: not determined. Information is from Jones *et al.* (1987) and Hough and Danson (1989). The list is not exhaustive.

PROPERTY	ARCHAEBACTERIA	EUBACTERIA	EUKARYOTES
1: Cell Size (Linear dimension)	About 1 micrometer	About 1 micrometer	About 10 micrometers
2: Cellular organelles	Absent	Absent	Present
3: Nuclear membrane	Absent	Absent	Present
4: Ribosomes (Subunit sizes)	30S, 50S	30S, 50S	40S, 60S
5: Length of 16S (18S) RNA	1500 nucleotides	1500 nucleotides	1800 nucleotides
6: Length of 23S (25-28S) RNA	2900 nucleotides	2900 nucleotides	3500 nucleotides +
7: mRNA binding site AUCACCUCC at 3' end of 16S (18S) RNA	Present	Present	Absent
8: Translation-elongation factor	Reacts with diphtheria toxin	No reaction with diphtheria toxin	Reacts with diphtheria toxin
9: Sensitive to chloramphenicol	No	Yes	No
10: Sensitive to anisomycin	Yes	No	Yes
11: Sensitive to kanamycin	No	Yes	No
12: Amino acid carried by initiator tRNA	Methionine	Formylmethionine	Methionine
13: Membrane lipids	Ether-linked, branched aliphatic chains	Ester-linked, straight aliphatic chains	Ester linked, straight aliphatic chains
14: tRNA - thymine in 'common' arm	Absent	Present in most tRNA's	Present in most tRNA's
15: tRNA - dihydrouracil	Absent (in all but 1 genus)	Present in all species	Present in all species
16: Cell wall	Variety of types: none incorporate muramic acid	Variety within 1 type: all incorporate muramic acid	No wall in animal cells, variety of types in other phyla

Properties 1-7 show similarities between archaebacteria and eubacteria. Properties 8-12 show similarities between archaebacteria and eukaryotes. Properties 13-16 are unique to the archaebacteria. Reproduced from Woese, 1981.

Table 1.2: Molecular traits of the three domains.

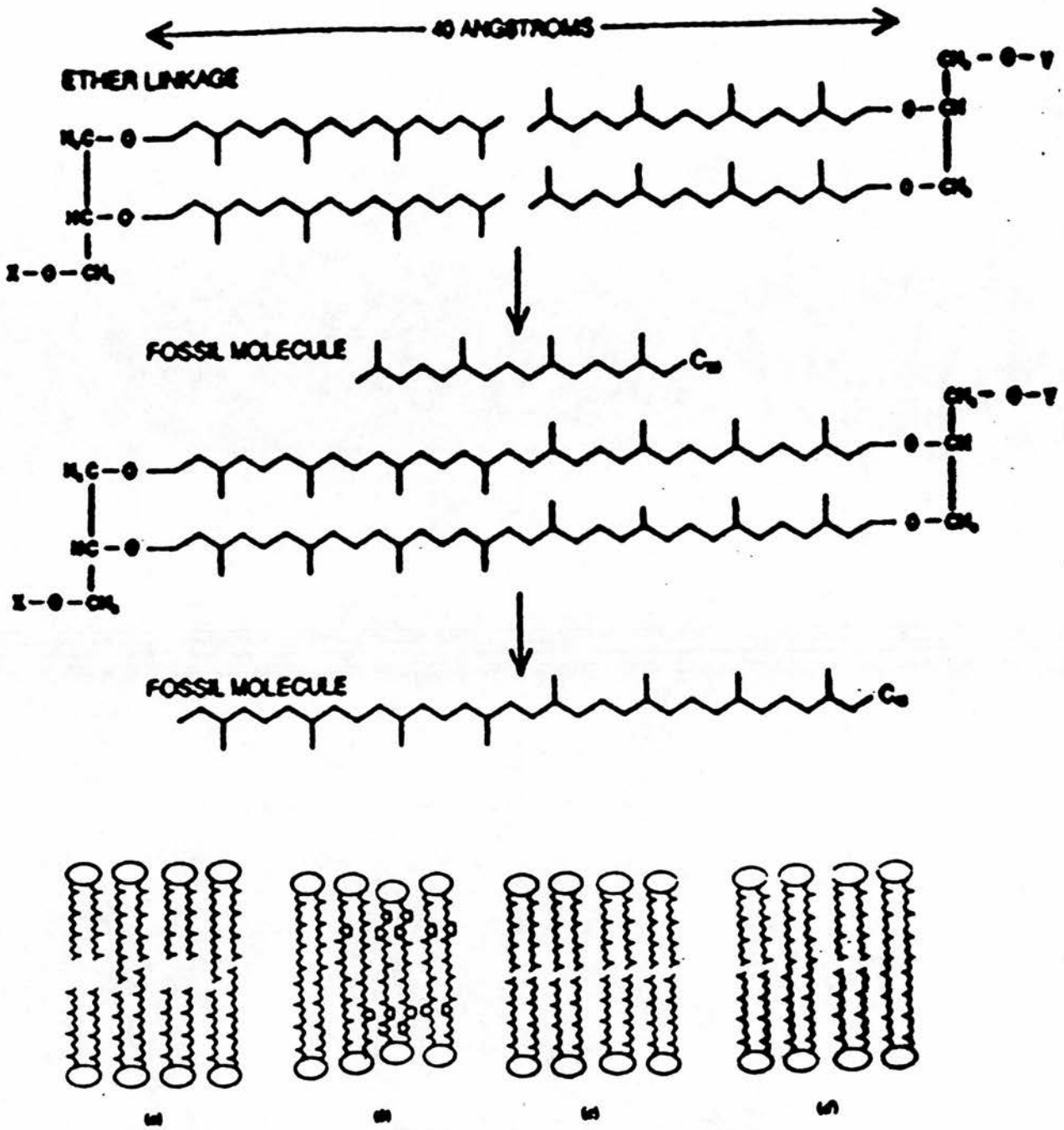


Figure 1.4: Membrane lipids of the archaebacteria. (i) shows lipids isolated from methanogenic archaebacteria. Worthy of note are the hydrocarbon tails which are branched rather than straight and the connection to the polar head group which is an ether rather than an ester linkage. (ii) shows proposed types of archaebacterial membranes: (a) 'zip' membrane formed by $C_{20}C_{25}$ and $C_{20}C_{20}$ diethers, (b) rigid monolayer of $C_{40}C_{40}$ tetraethers, (c) bilayer of $C_{20}C_{20}$ diethers, (d) mixed structure of $C_{20}C_{20}$ diethers and $C_{40}C_{40}$ tetraethers. Reproduced from Fewson (1986).

More recently, the peculiar physiology of the halophiles has suggested the development of other biotechnological applications. Halobacteria have been demonstrated to share sensitivity patterns with at least two major targets of cytostatic (anti-cancer) drugs: DNA topoisomerase II and cytoskeletal components (actin and tubulin). Therefore, the use of halobacteria as microbial pre-screens may well expedite the screening for new anti-cancer drugs (Sioud *et al.*, 1987). Other applications include the use of the unique retinal proteins of halobacteria as light biosensors, and of purple membrane as a reversible holographic medium for computer memories (Hampp *et al.*, 1990).

Halobacterial enzymes are a good raw material for enzyme technology primarily because of their high stability under conditions of high solute concentration (i.e. low water activity - they can catalyse reactions in organic media which contain less than 0.01% water). Such conditions are economically advantageous for several industrial processes. In addition, these enzymes often exhibit thermophilicity, enhanced stability at room temperature and allow reactions to occur using organic solvents. This latter property offers several benefits. (1) Enhanced stability: all reactions causing irreversible thermo-inactivation of enzymes involve free water. Such processes are, therefore, hindered by dehydration. For example, it has been found that in aqueous solution, chymotrypsin is irreversibly denatured after incubation at 60°C for a few minutes whereas it is stable for several hours at 100°C if it is dissolved in anhydrous octane (Klibanov, 1989). (2) Alterations in substrate specificities: the more hydrophobic the solvent, the lower is an enzyme's stereoselectivity. A number of peptides have been synthesised containing D-amino acid residues using subtilisin. This is impossible in water because of the enzymes strict L-stereoselectivity (Margolin *et al.*, 1987). (3) Molecular memory (activation by pre-treatment): if an enzyme is lyophilised from an aqueous solution in the presence of an activating ligand it will retain a structural 'imprint' of the ligand and thus remain highly active in organic media even in the absence of that ligand. This is due to the dehydrated enzyme displaying high conformational rigidity (Russell and Klibanov, 1988).

The production of unusual lipids by halophiles has the potential to be used for EOR (enhanced oil recovery) procedures (Post and Al-Harjan, 1988). This is a process by which crude oil retained in a deposit is forcibly extracted by injection of water. The water displaces the oil which is then pushed to the surface through the oil well. The process is much more efficient if the properties of water are altered, increasing its viscosity and reducing its surface tension. Halobacterial lipids have surfactant properties which are able to modify the properties of water for optimal EOR. Lastly, some halophiles produce industrially interesting biopolymers such as poly-3-hydroxybutyrate, a carbon storage material with the properties of a thermoplastic (with similar properties to polypropylene).

Thermophiles: The biotechnological interest in thermophilic archaebacteria is limited largely to the enzymes which they produce. Thermostable enzymes tend to be more stable with regards to other denaturing influences such as detergents and organic solvents resulting in the advantages outlined above. The only commercially successful application of a thermophilic enzyme to date is the use of *Thermus aquaticus* DNA polymerase (or, more recently, the DNA polymerase from the hyperthermophilic archaebacterium *Pyrococcus furiosus*) in polymerase chain reaction (PCR) technology. Further developments of such specialist applications are the most likely biotechnological future for thermostable enzymes (Cowan, 1992) - see Table 1.3.

Methanogens: Microbiologically produced methane can be used as an energy source and, on a small scale, this is already an industrial reality. Several countries, including India and China, have instituted projects to process agricultural and rural domestic wastes for methane (and hence energy) production (Hall and Hobson, 1988). Some methanogens also produce ethane, a product which may be of future interest in several ways: first, it is a higher value gas than methane (it has higher energy content and compressibility) and secondly, it can also be used for ethylene, and hence polymer, production.

enzyme	operating temperature	source	application
α-amylase	60-80	*B. acidocaldarius*	Starch processing Brewing/Baking Detergents Water treatment
β-galactosidase	75	*B. stearothermophilus*	Lactose hydrolysis Food processing
Protease	60-80	*T. vulgaris*	Baking/Brewing Food processing Detergents
Oxidoreductase	70-100	*C. acidophila*	Biosensors
DNA polymerase RNA polymerase	70-100	*T. aquaticus*	Molecular biology

Table 1.3: Applications of thermostable enzymes in biotechnology.

1.3 Glycolysis

1.3.1 The glycolytic pathway - general introduction

The pathways of sugar catabolism in eubacteria and eukaryotes have been extensively studied and, consequently, more information is available for the 15 enzymes associated with glycolysis (10 of which are commonly found amongst eubacterial and eukaryotic species) than for any comparable multi-enzyme system. Complete amino acid sequences are available for all of the glycolytic enzymes, and high resolution crystal structures have been derived for all of them (Fothergill-Gilmore and Michels, 1992). It is also known that some of the enzymes involved in the pathway have similar three-dimensional structures and that the sequences of the individual enzymes are strongly conserved throughout evolution. Among the 10 common glycolytic enzymes there are 18 recognisable domains (defining a domain to have at least four elements of regular secondary structure). Of these, three have the topology characteristic of nucleotide-binding domains and four are α/β-barrels. The remaining 11 all differ in detailed topology although many are superficially very similar. This substantial body of information has been exploited in several attempts to explain the evolution of the pathway.

There are four main theories which describe the ways in which enzymes may evolve (reviewed by Fothergill-Gilmore and Michels, 1992). The first suggests that consecutive enzymes in a metabolic pathway which bind the same substrate/product molecules may have evolved by a series of gene duplication events. A logical extension (and prediction) of this theory is that all of the enzymes in a pathway should be related to one another. The second theory suggests that enzymes which catalyse similar reactions (e.g. kinases) have developed as independent groups and have gradually specialised to act upon particular substrates by a process of divergent evolution. The third theory suggests that enzymes with a requirement for binding certain ligands (e.g. nucleotides) have diverged from a common ancestor. All of the above theories are based on the assumption that the evolution of enzymes and pathways is essentially divergent.

An alternative possibility, suggested by the fourth theory, is that similarities in enzyme structure and function can arise from convergent evolution. Metabolic pathways may, therefore, have evolved as chance associations of independently evolving enzymes which happened to act on similar substrate structures. The enzymes involved in the glycolytic pathway show very little sequence similarity to one another, despite the fact that certain tertiary structural elements (such as the α/β-barrel) occur frequently amongst them (Fothergill-Gilmore and Michels, 1992). It is, therefore, impossible to say which of the above four theories best explains the evolution of the enzymes in the pathway. Given the evidence that thermophilic archaebacteria contain only two of the common glycolytic enzymes it is tempting to speculate that the pathway, as a whole, evolved from the 'bottom up' i.e. pyruvate kinase and enolase are the progenitor enzymes for the rest of the enzymes of the pathway.

1.3.2 The glycolytic pathway - (a) eukaryotes and eubacteria

There are two major generally recognised pathways for glycolysis. The Embden-Meyerhof pathway is characteristic of eukaryotes and a number of anaerobic and facultatively anaerobic eubacteria. In this pathway, glucose is phosphorylated by hexokinase and converted to fructose 1,6-*bis*phosphate by phosphoglucose isomerase. The fructose 1,6-*bis*phosphate then suffers an aldol cleavage, catalysed by aldolase, to two interconvertible triose phosphates: dihydroxyacetone phosphate and glyceraldeyde 3-phosphate (they are interconverted by triose phosphate isomerase). These then enter a common catabolic route to pyruvate - the 'trunk pathway' (Fig. 1.5): the trunk pathway enzymes are, therefore, glyceraldehyde 3-phosphate dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase, enolase and pyruvate kinase. In the complete Embden-Meyerhof pathway there is a net gain in energy of two molecules of ATP (-2, +4). Historically, this low level of ATP generation was thought to be an indication of the antiquity of the pathway. The subsequent elucidation of other glycolytic pathways has refuted the suggestion that Embden-Meyerhof glycolysis may be the most ancient of

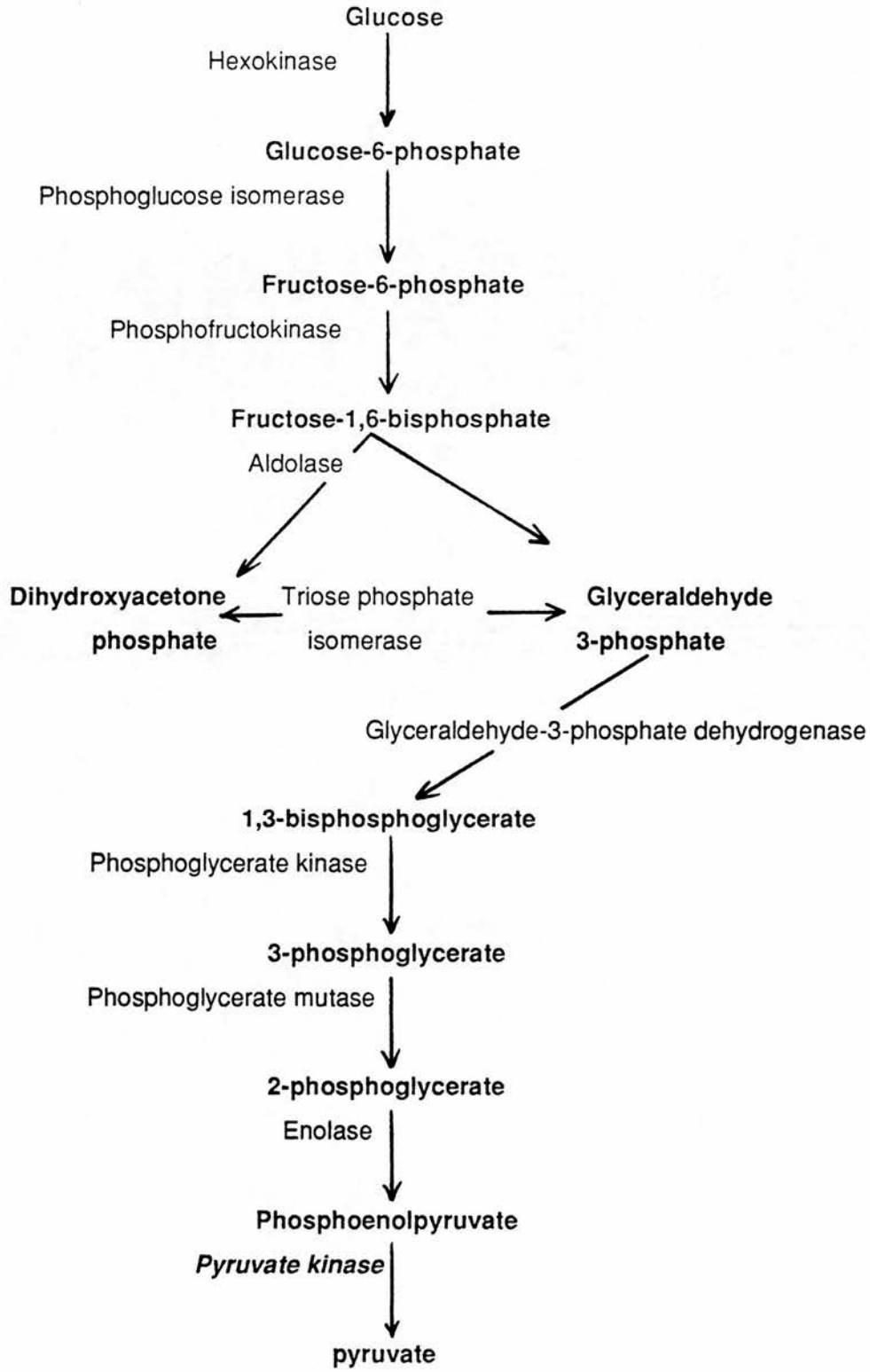


Figure 1.5. The Embden-Meyerhof pathway of glycolysis. The reaction catalysed by pyruvate kinase is highlighted.

metabolic pathways.

In many strictly aerobic eubacteria, the key enzyme 6-phosphofructokinase is absent and hence the glycolytic pathway described above cannot operate. In these organisms, glucose is instead catabolised by the Entner-Doudoroff pathway. Glucose 6-phosphate is oxidised to 6-phosphogluconate which is then dehydrated to 2-keto-3-deoxy-6-phosphogluconate. This then undergoes aldol cleavage to glyceraldehyde 3-phosphate (which enters the trunk pathway) and pyruvate. In this pathway, there is a net gain in energy of only one molecule of ATP (-1, +2).

1.3.3 The glycolytic pathway - (b) archaebacteria

The true central pathway of sugar catabolism is, therefore, the trunk sequence from triose phosphate to pyruvate (Cooper, 1986). Since the 'discovery' of the archaebacteria, it has been a priority to discover if the same pathway also exists in these organisms. Only if this is the case can archaebacterial enzymes be selected for comparison with their counterparts from other organisms (i.e the enzymes are found in comparable metabolic contexts). Many of the extreme halophiles use proteins and amino acids, rather than carbohydrates, as a source of carbon (Larsen, 1981). *Halobacterium saccharovorum*, however, catabolises glucose and galactose via a modified Entner-Doudoroff pathway (Fig. 1.6; Tomlinson *et al.*, 1974). Glucose 6-phosphate and 6-phosphogluconate are not oxidised by this organism, suggesting that neither the traditional Entner-Doudoroff nor the Embden-Meyerhof pathway is operating. Instead, glucose is oxidised to gluconate by an NAD-dependent glucose dehydrogenase and is then converted stoichiometrically into 2-keto-3-deoxygluconate by gluconate dehydratase. This is phosphorylated by ATP and the 2-keto-3-deoxy-6-phosphogluconate so generated undergoes aldol cleavage to glyceraldehyde 3-phosphate and pyruvate. Activities of the enzymes of the common trunk sequence discussed above have been detected in *H. saccharovorum* and hence it is likely that the glyceraldehyde 3-phosphate is metabolised to pyruvate in the same manner as it is in eukaryotes and eubacteria. The net energy yield of the pathway is again

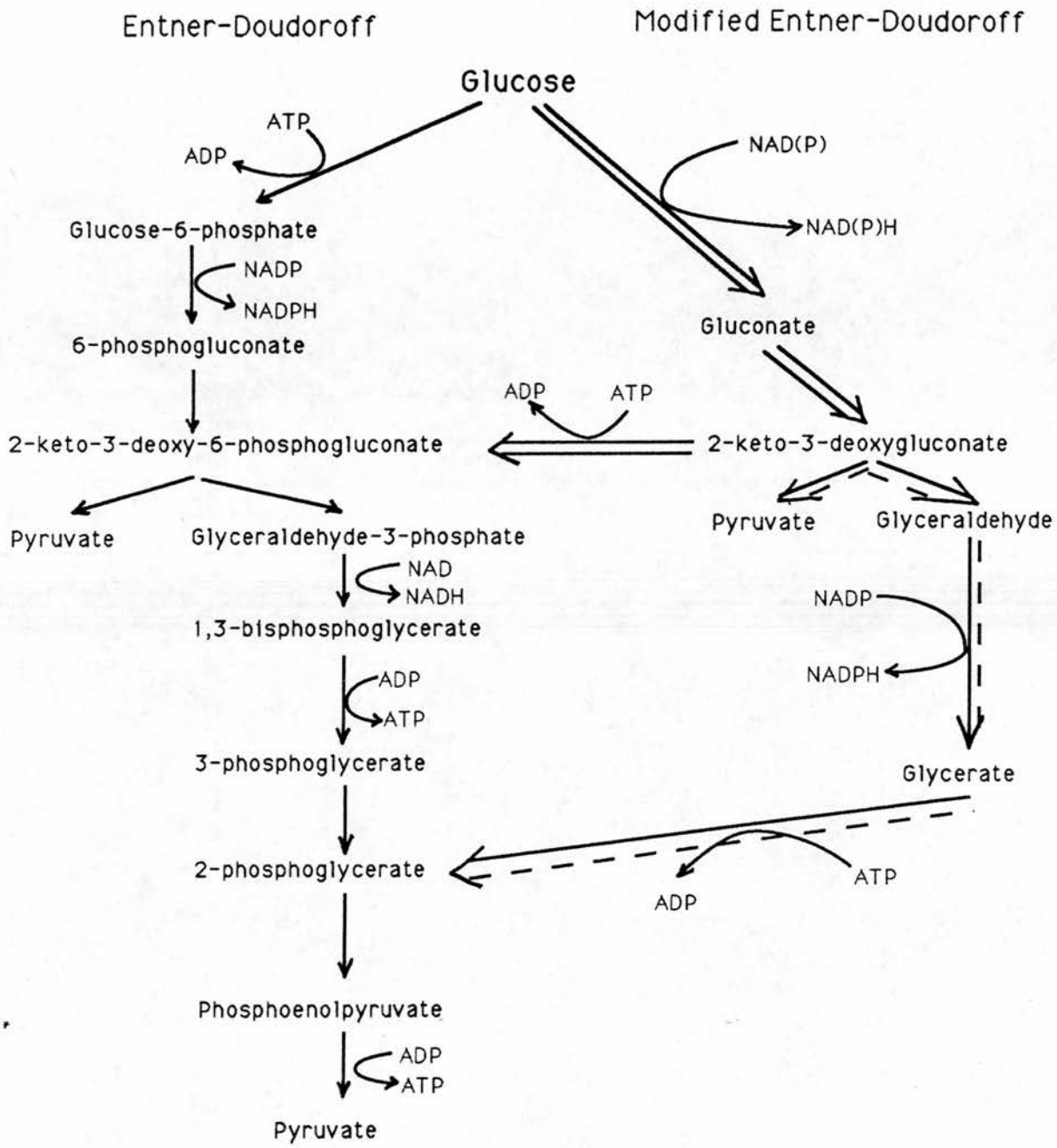


Figure 1.6: The modified Entner-Doudoroff pathways utilised by halophilic and thermophilic archaebacterial species. Halophiles: double line, thermophiles: dashed line

one ATP (-1, +2). The modified pathway of *H. saccharovorum* is not unique to the archaebacteria; it has also been reported in species of *Clostridium* and other eubacteria (Andreesen and Gottschalk, 1969). In these organisms, however, the pathway operates in gluconate-grown cells - there have been no reports on whether glucose can be catabolised.

In thermoacidophilic archaebacteria, another more drastic modification of the Entner-Doudoroff pathway is employed (Fig. 1.6). In *Sulfolobus solfataricus* (DeRosa *et al.*, 1984), glucose is converted to gluconate via an NAD(P)-dependent dehydrogenation step. The gluconate is then dehydrated to 2-keto-3-deoxygluconate which is then cleaved directly to pyruvate and glyceraldehyde. In *Thermoplasma acidophilum* it has been demonstrated that the glyceraldehyde is oxidised to glycerate by an NADP-dependent dehydrogenase and the glycerate is then converted to 2-phosphoglycerate by glycerate kinase. Enolase and pyruvate kinase complete the pathway to give the second molecule of pyruvate (Budgen and Danson, 1986). This discovery is noteworthy in two respects. First, this pathway has a net energy generation of zero ATP (-1, +1) and hence *T. acidophilum* must use other reactions for energy generation (in fact, acetyl CoA generated from pyruvate via pyruvate:ferredoxin oxidoreductase is converted to acetate with the concomitant production of ATP, Budgen and Danson, unpublished observations). Secondly, in the thermoacidophilic archaebacteria, the common trunk sequence of enzymes is largely absent - only enolase and pyruvate kinase remain. This refutes the idea that this pathway is ubiquitous throughout the phylogenetic tree of life and casts doubt on the proposal that it represents the most ancient of metabolic pathways (Gest and Schopf, 1983).

Methanogenic archaebacteria are mostly autotrophs, fixing carbon eventually into acetyl CoA and hence hexose metabolism is mainly in the direction of carbohydrate synthesis. In these organisms, the synthesis of glucose from acetyl CoA is effected by a reversal of the traditional Embden-Meyerhof pathway. Interestingly, the pathways of gluconeogenesis have not been established in thermoacidophilic archaebacteria - perhaps a reversal of the non-phosphorylated Entner-Doudoroff pathway could occur but this proposition awaits investigation.

A summary of the evidence presented above, plus information about the citric acid cycle present in the archaebacteria can be seen in Table 1.4.

1.4 Pyruvate kinase

1.4.1 General introduction and sequences

Pyruvate kinase (ATP:pyruvate 2-O-phosphotransferase, E.C. 2.7.1.40) catalyses the essentially irreversible reaction:

$$\text{phosphoenolpyruvate} + \text{MgADP} + H^{+} \xrightarrow{Mg^{2+}\ K^{+}} \text{pyruvate} + \text{MgATP}$$

The enzyme has an absolute requirement for two divalent cations (Mg^{2+}) and one monovalent cation (K^{+}) as cofactors. One of the magnesium ions is found in close association with the protein whilst the other is complexed with the nucleotide (Muirhead, 1987). The enzyme has a low nucleotide specificity in that multiple phosphodonors can be utilised: UTP, GTP, CTP, ITP and dADP (Strominger, 1955). It catalyses the second ATP-forming reaction of glycolysis (or the only one in the archaebacterial non-phosphorylated Entner-Doudoroff pathway), and is a major control point of the pathway as a whole - under physiological conditions the reaction is irreversible in the favour of ATP formation. The reaction mechanism involves the addition of a proton and a direct in-line transfer of the phospho group. The product, pyruvate, is the first non-phosphorylated intermediate in the eukaryotic pathway and occupies a central role in metabolism.

Pyruvate kinase is one of only two glycolytic enzymes still presumed to be ubiquitous, the other being enolase. It has been purified from many sources both eukaryotic and eubacterial; this has recently been reviewed (Guderley *et al.*, 1989). Generally, the enzyme is a homotetramer of relative molecular mass (Mr) 250000 although the subunit size varies between species within the Mr range 55-64000. In two cases, however,

organism	glucose catabolism	gluconeogenesis	citric acid cycle
HALOPHILES:			
H. saccharovorum	Modified Entner-Doudoroff pathway	Reverse Embden-Meyerhof pathway	Complete oxidative cycle
THERMOPHILES:			
Sulfolobus	Non-phosphorylated Entner-Doudoroff pathway	Unknown	Autotrophy - reductive cycle; Heterotrophy - oxidative cycle
Thermoplasma	Non-phosphorylated Entner-Doudoroff pathway	Unknown	Complete oxidative cycle
METHANOGENS:			
Methanobacterium	Entner-Doudoroff pathway	Reverse Embden-Meyerhof pathway	Incomplete reductive cycle

Table 1.4: Central metabolic pathways of the archaebacteria.

the enzyme has a native Mr equivalent to half of the normal value and it may be speculated that these are dimeric forms of the enzyme: *Zymomonas mobilis*, (Pawluk *et al.*, 1988) and *Schizosaccharomyces pombe* (Duncan *et al.*, 1989). Thirteen pyruvate kinase sequences are now known and an alignment of these, generated using the multiple sequence alignment program CLUSTAL (Higgins and Sharp, 1988), is shown in Fig. 1.7.

1.4.2 Regulation of pyruvate kinase

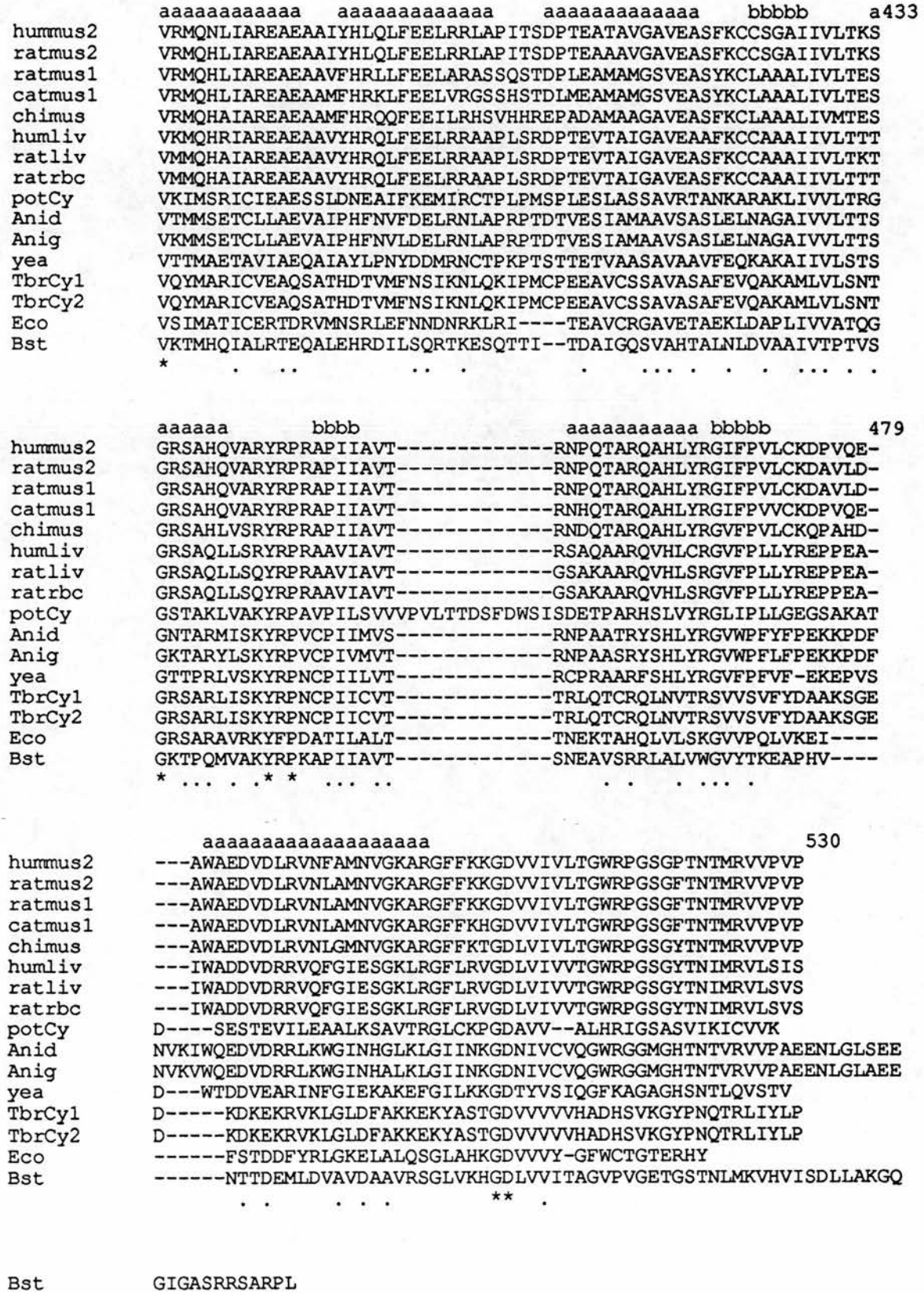
Pyruvate kinases are subject to both coarse and fine control. For example, yeast pyruvate kinase is subject to coarse control by phosphorylation by cAMP-dependent protein kinase *in vitro* (Blair and Harman, 1986). Sequencing studies have revealed a potential cAMP-dependent protein kinase binding site (Burke *et al.*, 1983) and the phosphorylation at this site appears to activate the enzyme. Unfortunately, results have proved variable and hence, the physiological significance of this observation has yet to be determined. The yeast enzyme is subject to further, fine control by allosteric effectors. It is activated by fructose 1,6-*bis*phosphate and inhibited by ATP, gluconeogenic amino acids, citrate and high pH (reviewed by McNally, 1991).

Pyruvate kinase can be isolated from mammalian tissues as four isoenzymes which have different kinetic properties, reflecting the different metabolic requirements of the tissues of origin. The M1 isoenzyme is found in skeletal muscle and displays predominantly hyperbolic Michaelis-Menten kinetics. The other isoenzymes (M2 in kidney, adipose tissue and lung; L in liver; and R in erythrocytes) all show sigmoidal kinetics with respect to PEP and are allosterically regulated (Hall and Cottam, 1978). The enzyme from yeast has properties similar to those of the M2 isoenzyme (Hunsley and Suelter, 1969): a variety of molecules are known to act as allosteric effectors and, in addition, the liver enzyme can be regulated by phosphorylation. In this isoenzyme, phosphorylation of a serine residue near the amino-terminus

Figure 1.7: Sequence alignment of pyruvate kinase sequences generated by the multiple sequence alignment program CLUSTAL (Higgins and Sharp, 1988). Regions of secondary structure are indicated with a's (for α-helices) and b's (for β-strands). Identical residues are marked with an * and conserved residues with a period.

```
hummus2                                                    SKPHSEAGTAFIQTQQLH
ratmus2                                                    PKPDSEAGTAFIQTQQLH
ratmus1                                                    PKPDSEAGTAFIQTQQLH
catmus1                                                    SKPHSDVGTAFIQTQQLH
chimus                                                     SKHHDAGTAFIQTQQLH
humliv                                           EGPAGYLRRASVAQLTQELGTAFFQQQQLP
ratliv                                           EGPAGYLRRASVAQLTQELGTAFFQQQQLP
ratrbc      SVQENTLPQQLWPWIFRSQKDLAKSALSGAGGPAGYLRRASVAQLTQELGTAFFQQQQLP
Anid                                                                   AASSSLD
Anig                                                                   AASSSLD


                 aaaaaaaa                bbbbbbbb     aaaaaaaaaaa  bbbbbb  77
hummus2     AAMADTFLEHMCRLDIDSPPITA-RNTGIICTIGPASRSVETLKEMIKSGMNVARLNFSH
ratmus2     AAMADTFLEHMCRLDIDSAPITA-RNTGIICTIGPASRSVEMLKEMIKSGMNVARLNFSH
ratmus1     AAMADTFLEHMCRLDIDSAPITA-RNTGIICTIGPASRSVEMLKEMIKSGMNVARLNFSH
catmus1     AAMADTFLEHMCRLDIDSPPITA-RNTGIICTIGPASRSVEILKEMIKSGMNVARLNFSH
chimus      AAMADTFLEHMCRLDIDSEPTIA-RNTGIICTIGPASRSVDKLKEMIKSGMNVARLNFSH
humliv      AAMADTFLEHLCLLDIDSEPVAA-RSTSIIATIGPASRSVGRLKEMIKAGMNIARLNFSH
ratliv      AAMADTFLEHLCLLDIDSQPVAA-RSTSIIATIGPASRSVDRLKEMIKAGMNIARLNFSH
ratrbc      AAMADTFLEHLCLLDIDSEPVAA-RSTSIIATIGPASRSVDRLKEMIKAGMNIARLNFSH
potCy             ANIDIAGIMKDLPNDGRIPKTKIVCTLGPSSRTVPMLEKLLRAGMNVARFNFSH
Anid        HLSNRMKLEWHSKLNTEMVPAKNFRRTSIICTIGPKTNSVEKINALRRAGLNVVRMNFSH
Anig        HLSNRMKLEWHSKLNTEMVPSKNFRRTSIIGTIGPKTNSVEKINSLRTAGLNVVRMNFSH
yea              SRLERLTSLN--VVAGSDLRRTSIIGTIGPKTNNPETLVALRKAGLNIVRMNFSH
TbrCy1           SQLEHNIGLSIFEPVAKH-RANRIVCTIGPSTQSVEALKNLMKSGMSVARMNFSH
TbrCy2           SQLEHNIGLSIFEPVAKH-RANRIVCTIGPSTQSVEALKNLMKSGMSVARMNFSH
Eco                            KKTKIVCTIGPKTESEEMLAKMLDAGMNVMRLNFSH
Bst                           KRKTKIVCTIGPASESVDKLVQLMEAGMNVARLNFSH
                                 . *. *.** . .   . .  .*... *.****


               aaaaaaaaaaaaaaaaaaa       bbbbbbb    bbbbbbb bbbb bbbb bbb137
hummus2     GTHEYHAETIKNVRTATESFASDPILYRPVAVALDTKGPEIRTGLIKGSGTAEVELKKGA
ratmus2     GTHEYHAETIKNVRAATESFASDPILYRPVAVALDTKGPEIRTGLIKGSGTAEVELKKGA
ratmus1     GTHEYHAETIKNVRAATESFASDPILYRPVAVALDTKGPEIRTGLIKGSGTAEVELKKGA
catmus1     GTHEYHAETIKNVRAATESFASDPIRYRPVAVALDTKGPEIRTGLIKGSGTAEVELKKGA
chimus      GTHEYHEGTIKNVREATESFASDPITYRPVAIALDTKGPEIRTGLIKGSGTAEVELKKGA
humliv      GSHEYHAETIANVREAVESFAGSPLSYRPVAIALDTKGPEIRTGILQGGPESEVELVKGS
ratliv      GSHEYHAESIANIREATESFATSPLSYRPVAIALDTKGPEIRTGVLQGGPESEVEIVKGS
ratrbc      GSHEYHAESIANIREATESFATSPLSYRPVAIALDTKGPEIRTGVLQGGPESEVEIVKGS
potCy       GTHEYGQETLDNLKIAMQNTQIL------CAVMLDTKGPEIRTGFLTDGK--PIQLKEGQ
Anid        GSYEYHQSVIDHAREAEKQAAG-----RPVAIALDTKGPEIRTGNTVGDK--DIPIKAGH
Anig        GSYEYHQSVIDNAREAAKTQVG-----RPLAIALDTKGPEIRTGNTPDDK--DIPIKQGH
yea         GSYEYHKSVIDNARKSEELYPG-----RPLAIALDTKGPEIRTGTTTNDV--DYPIPPNH
TbrCy1      GSHEYHQTTINNVRAAAAELGLH------IGIALDTKGPEIRTGLFKDG---EVSFAPGD
TbrCy2      GSHEYHQTTINNVRAAAAELGLH------IGIALDTKGPEIRTGLFKDG---EVTFAPGD
Eco         GDYAEHGQRIQNLRNVMSKTGKT------AAILLDTKGPEIRTMKLEGGN--DVSLKAGQ
Bst         GDHEEHGRRIANIREAAKRTGRT------VAILLDTKGPEIRTHNMENGA---IELKEGS
            *...     . . .                .. **********    .       .  .


            b aaaaaa   aaaaaa      aaaaaaaaaaaaa  bbbbbbb  bbbbbbb      195
hummus2     TLKITLDNAYMEKCDENILWLDYKNICKVVEVGSKIYVDDGLISLQVKQKGADF--LVTE
ratmus2     TLKITLDNAYMEKCDENILWLDYKNICKVVEVGSKIYVDDGLISLQVKEKGADY--LVTE
ratmus1     TLKITLDNAYMEKCDENILWLDYKNICKVVEVGSKIYVDDGLISLQVKEKGADY--LVTE
catmus1     TLKITLDNAYMEKCDENVLWLDYKNICKVVEVGSKVYVDDGLISLLVKEKGADF--LVTE
chimus      ALKVTLDNAFMENCDENVLWVDYKNLIKVIDVGSKIYVDDGLISLLVKEKGKDF--VMTE
humliv      QVLVTVDPAFRTRGNANTVWVDYPNIVRVVPVGGRIYIDDGLISLVVQKISPEG--LVTQ
ratliv      QVLVTVDPKFQTRGDAKTVWVDYHNITRVVAVGGRIYIDDGLISLVVQKIGPEG--LVTE
ratrbc      QVLVTVDPAFQTRGDAKTVWVDYHNITRVVAVGGRIYIDDGLISLVVQKIGPEG--LVTE
potCy       EITVSTDYTI--KGNEEMISMSYKKLVMDLKPGNTILCADGTITLTVLSCDPPSGTVRCR
Anid        EMNISTDEQYATASDDQNMYVDYKNITKVISAGKLIYVDDGILSFEVLEVVDDK-TLRVR
Anig        ELNITTDEQYATASDDKNMYLDYKNITKVISPGKLIYVDDGILSFEVLEVVDDK-TIRVR
```

```
yea          EMIFTTDDKYAKACDDKIMYVDYKNITKVISAGRIIYVDDGVLSFQVLEVVDDK-TLKVK
TbrCy1       IVCVTTDPAYEKVGTKEKFYIDYPQLTNAVRPGGSIYVDDGVMTLRVVSKEDDR-TLKCH
TbrCy2       IVCVTTDPAYEKVGTKEKFYIDYPQLTKAVPVGGSIYVDDGVMTLRVLSKEDDR-TLKCH
Eco          TFTFTTDKSV--IGNSEMVAVTYEGFTTDLSVGNTVLVDDGLIGMEVTAIEGNK--VICK
Bst          KLVISMSEV---LGTPEKISVTYPSLIDDVSVGAKILLDDGLISLEVNAVDKQAGEIVTT
                 . .       . .   ..*  .   .  *  .  .** ... *         .

                     bbbbb bbb  bbbb     bbb   aaaaaaaa   bbbb     aaaaaaaaaa254
hummus2      VENGGSLGSKKGVNLPGAAVDLPAVSEKDIQDL-KFGVEQDVDMVFASFIRKASDVHEVR
ratmus2      VENGGSLGSKKGVNLPGAAVDLPAVSEKDIQDL-KFGVEQDVDMVFASFIRKAADVHEVR
ratmus1      VENGGSLGSKKGVNLPGAAVDLPAVSEKDIQDL-KFGVEQDVDMVFASFIRKAADVHEVR
catmus1      VENGGSLGSKKGVNLPGAAVDLPAVSEKDIQDL-KFGVEQDVDMVFASFIRKASDVHEVR
chimus       VENGGMLGSKKGVNLPGAAVDLPAVSEKDIQDL-KFGVEQNVDMVFASFIRKAADVHAVR
humliv       VENGGVLGSRKGVNLPGAQVDLPGLSEQDVRDL-RFGVEHGVDIVFASFVRKASDVAAVR
ratliv       VEHGGILGSRKGVNLPNTEVDLPGLSEQDLLDL-RFGVQHNVDIIFASFVRKASDVLAVR
ratrbc       VENGGILGSRKGVNLPNTEVDLPGLSEQDLLDL-RFGVQHNVDIIFASFVRKASDVLAVR
potCy        CENTATLGERKNVNLPGVVVDLPTLTEKDKEDILEWGVPNNIDMIALSFVRKGSDLVNVR
Anid         CLNNGNISSRKGVNLPGTDVDLPALSEKDISDL-KFGVKNKVDMVLASFIRRGSDIRHIR
Anig         CLNNGNISSRKGVNLPGTDVDLPALSEKDIADL-KFGVRNKVDMVFASFIRRGSDIRHIR
yea          ALNAGKICSHKGVNLPGTDVDLPALSEKDKEDL-RFGVKNGVHMVFASFIRTANDVLTIR
TbrCy1       VNNHHRLTDRRGINLPGCEVDLPAVSEKDRKDL-EFGVAQGVDMIFASFIRTAEQVREVR
TbrCy2       VNNHHRLTDRRGINLPGCEVDLPAVSEKDRKDL-EFGVAQGVDMIFASFIRTAEQVREVR
Eco          VLNNGDLGENKGVNLPGVSIALPALAEKDKQDLI-FGCEQGVDFVAASFIRKRSDVIEIR
Bst          VLNGGVLKNKKGVNVPGVKVNLPGITEKDRADIL-FGIRQGIDFIAASFVRRASDVLEIR
               .   . .....*.*.  ..**...*.*  *.  .*  . ....  **.*  ...  .*

             aa       bbbbbb  aaaaa aaaaaa    bbbb  aaaaa   aaaaaaaaaaaaa313
hummus2      KVLGEK-GKNIKIISKIENHEGVRRFDEILEASDGIMVARGDLGIEIPAEKVFLAQKMMI
ratmus2      KVLGEK-GKNIKIISKIENHEGVRRFDEILEASDGIMVARGDLGIEIPAEKVFLAQKMMI
ratmus1      KVLGEK-GKNIKIISKIENHEGVRRFDEILEASDGIMVARGDLGIEIPAEKVFLAQKMMI
catmus1      KVLGEK-GKNIKIISKIENHEGVRRFDEILEASDGIMVARGDLGIEIPAEKVFLAQKMMI
chimus       KVLGEK-GKHIKIISKIENHEGVRRFDEIMEASDGIMVARGDLGIEIPAEKVFLAQKMMI
humliv       AALGPE-GHGIKIISKIENHEGVKRFDEILEVSDGIMVARGDLGIEIPAEKVFLAQKMMI
ratliv       DALGPE-GQNIKIISKIENHEGVKKFDEILEVSDGIMVARGDLGIEIPAEKVFLAQKMMI
ratrbc       DALGPE-GQNIKIISKIENHEGVKKFDEILEVSDGIMVARGDLGIEIPAEKVFLAQKMMI
potCy        KVLGPH-AKRIQLMSKVENQEGVINFDEILRETDSFMVARGDLGMEIPVEKIFLAQKMMI
Anid         EVLGEE-GREIQIIAKIENQQGVNNFDEILEETDGVMVARGDLGIEIPAPKVFIAQKMMI
Anig         EVLGEE-GREIQIIAKIENQQGVNNFDEILEETDGVMVARGDLGIEIPAPKVFIAQKMMI
yea          EVLGEQ-GKDVKIIVKIENQQGVNNFDEILKVTDGVMVARGDLGIEIPAPEVLAVQKKLI
TbrCy1       AALGEK-GKDILIISKIENHQGVQNIDSIIEASNGIMVARGDLGVEIPAEKVCVAQMCII
TbrCy2       AALGEK-GKDILIISKIENHQGVQNIDSIIEASNGIMVARGDLGVEIPAEKVCVAQMCII
Eco          EHLKAHGGENIHIISKIENQEGLNNFDEILEASDGIMVARGDLGVEIPVEEVIFAQKMMI
Bst          ELLEAHDALHIQIIAKIENEEGVANIDEILEAADGLMVARGDLGVEIPAEEVPLIQKLLI
               *  . .  . .. *.**..*. ..*.*.  ... ********.***. ..   *. .*

             aaaa  bbbbbb            aaaaaaaaaaaa    bbbbbb        aaaaaa373
hummus2      GRCNRAGKPVICATQMLESMIKKPRPTRAEGSDVANAVLDGADCIMLSGETAKGDYPLEA
ratmus2      GRCNRAGKPVICATQMLESMIKKPRPTRAEGSDVANAVLDGADCIMLSGETAKGDYPLEA
ratmus1      GRCNRAGKPVICATQMLESMIKKPRPTRAEGSDVANAVLDGADCIMLSGETAKGDYPLEA
catmus1      GRCNRAGKPVICATQMLESMIKKPRPTRAEGSDVANAVLDGADCIMLSGETAKGDYPLEA
chimus       GRCNRAGKPIICATQMLESMIKKPRPTRAEGSDVANAVLDGADCIMLSGETAKGDYPLEA
humliv       GRCNLAGKPVVCATQMLESMITKPRPTRAETSDVANAVLDGADCIMLSGETAKGNFPVEA
ratliv       GRCNLAGKPVVCATQMLESMITKARPTRAETSDVANAVLDGADCIMLSGETAKGSFPVEA
ratrbc       GRCNLAGKPVVCATQMLESMITKARPTRAETSDVANAVLDGADCIMLSGETAKGNFPVEA
potCy        YKCNLAGKAVVTATQMLESMIKSPAPTRAEATDVANAVLDGTDCVMLSGESAAGAYPELA
Anid         AKCNIKGKPVICATQMLESMTYNPRPTRAEVSDVANAVLDGADCVMLSGETAKGNYPCEA
Anig         AKCNIKGKPVICATQMLESMTYNPRPTRAEVSDVANAVLDGADCVMLSGETAKGNYPNEA
yea          AKSNLAGKPVICATQMLESMTYNPRPTRAEVSDVGNAILDGADCVMLSGETAKGNYPINA
TbrCy1       SKCNVVGKPVICATQMLESMTSNPRPTRAEVSDVANAVLNGADCVMLSGETAKGKYPNEV
TbrCy2       SKCNVVGKPVICATQMLESMTSNPRPTRAEVSDVANAVLNGADCVMLSGETAKGKYPNEV
Eco          EKCIRARKVVITATQMLDSMIKNPRPTRAEAGDVANAILDGTDAVMLSGESAKGKYPLEA
Bst          KKCNMLGKPVITATQMLDSMQRNPRPTRAEASDVANAIFDGTDAVMLSGETAAGQYPVEA
               ..    * .. *****.**  .. ***** .**.**...*.* .*****.* * .*  .
```


aaaaaaaaaaaa aaaaaaaaaaaaa aaaaaaaaaaaaa bbbbb a433
hummus2 VRMQNLIAREAEAAIYHLQLFEELRRLAPITSDPTEATAVGAVEASFKCCSGAIIVLTKS
ratmus2 VRMQHLIAREAEAAIYHLQLFEELRRLAPITSDPTEAAAVGAVEASFKCCSGAIIVLTKS
ratmus1 VRMQHLIAREAEAAVFHRLLFEELARASSQSTDPLEAMAMGSVEASYKCLAAALIVLTES
catmus1 VRMQHLIAREAEAAMFHRKLFEELVRGSSHSTDLMEAMAMGSVEASYKCLAAALIVLTES
chimus VRMQHAIAREAEAAMFHRQQFEEILRHSVHHREPADAMAAGAVEASFKCLAAALIVMTES
humliv VKMQHRIAREAEAAVYHRQLFEELRRAAPLSRDPTEVTAIGAVEAAFKCCAAAIIVLTTT
ratliv VMMQHAIAREAEAAVYHRQLFEELRRAAPLSRDPTEVTAIGAVEASFKCCAAAIIVLTKT
ratrbc VMMQHAIAREAEAAVYHRQLFEELRRAAPLSRDPTEVTAIGAVEASFKCCAAAIIVLTTT
potCy VKIMSRICIEAESSLDNEAIFKEMIRCTPLPMSPLESLASSAVRTANKARAKLIVVLTRG
Anid VTMMSETCLLAEVAIPHFNVFDELRNLAPRPTDTVESIAMAAVSASLELNAGAIVVLTTS
Anig VKMMSETCLLAEVAIPHFNVLDELRNLAPRPTDTVESIAMAAVSASLELNAGAIVVLTTS
yea VTTMAETAVIAEQAIAYLPNYDDMRNCTPKPTSTTETVAASAVAAVFEQKAKAIIVLSTS
TbrCy1 VQYMARICVEAQSATHDTVMFNSIKNLQKIPMCPEEAVCSSAVASAFEVQAKAMLVLSNT
TbrCy2 VQYMARICVEAQSATHDTVMFNSIKNLQKIPMCPEEAVCSSAVASAFEVQAKAMLVLSNT
Eco VSIMATICERTDRVMNSRLEFNNDNRKLRI----TEAVCRGAVETAEKLDAPLIVVATQG
Bst VKTMHQIALRTEQALEHRDILSQRTKESQTTI--TDAIGQSVAHTALNLDVAAIVTPTVS
aaaaaa bbbb aaaaaaaaaaa bbbbb 479
hummus2 GRSAHQVARYRPRAPIIAVT----------RNPQTARQAHLYRGIFPVLCKDPVQE-
ratmus2 GRSAHQVARYRPRAPIIAVT----------RNPQTARQAHLYRGIFPVLCKDAVLD-
ratmus1 GRSAHQVARYRPRAPIIAVT----------RNPQTARQAHLYRGIFPVLCKDAVLD-
catmus1 GRSAHQVARYRPRAPIIAVT----------RNHQTARQAHLYRGIFPVVCKDPVQE-
chimus GRSAHLVSRYRPRAPIIAVT----------RNDQTARQAHLYRGVFPVLCKQPAHD-
humliv GRSAQLLSRYRPRAAVIAVT----------RSAQAARQVHLCRGVFPLLYREPPEA-
ratliv GRSAQLLSQYRPRAAVIAVT----------GSAKAARQVHLSRGVFPLLYREPPEA-
ratrbc GRSAQLLSQYRPRAAVIAVT----------GSAKAARQVHLSRGVFPLLYREPPEA-
potCy GSTAKLVAKYRPAVPILSVVVPVLTTDSFDWSISDETPARHSLVYRGLIPLLGEGSAKAT
Anid GNTARMISKYRPVCPIIMVS----------RNPAATRYSHLYRGVWPFYFPEKKPDF
Anig GKTARYLSKYRPVCPIVMVT----------RNPAASRYSHLYRGVWPFLFPEKKPDF
yea GTTPRLVSKYRPNCPIILVT----------RCPRAARFSHLYRGVFPFVF-EKEPVS
TbrCy1 GRSARLISKYRPNCPIICVT----------TRLQTCRQLNVTRSVVSVFYDAAKSGE
TbrCy2 GRSARLISKYRPNCPIICVT----------TRLQTCRQLNVTRSVVSVFYDAAKSGE
Eco GRSARAVRKYFPDATILALT----------TNEKTAHQLVLSKGVVPQLVKEI----
Bst GKTPQMVAKYRPKAPIIAVT----------SNEAVSRRLALVWGVYTKEAPHV----
aaaaaaaaaaaaaaaaaa 530
hummus2 ---AWAEDVDLRVNFAMNVGKARGFFKKGDVVIVLTGWRPGSGPTNTMRVVPVP
ratmus2 ---AWAEDVDLRVNLAMNVGKARGFFKKGDVVIVLTGWRPGSGFTNTMRVVPVP
ratmus1 ---AWAEDVDLRVNLAMNVGKARGFFKKGDVVIVLTGWRPGSGFTNTMRVVPVP
catmus1 ---AWAEDVDLRVNLAMNVGKARGFFKHGDVVIVLTGWRPGSGFTNTMRVVPVP
chimus ---AWAEDVDLRVNLGMNVGKARGFFKTGDLVIVLTGWRPGSGYTNTMRVVPVP
humliv ---IWADDVDRRVQFGIESGKLRGFLRVGDLVIVVTGWRPGSGYTNIMRVLSIS
ratliv ---IWADDVDRRVQFGIESGKLRGFLRVGDLVIVVTGWRPGSGYTNIMRVLSVS
ratrbc ---IWADDVDRRVQFGIESGKLRGFLRVGDLVIVVTGWRPGSGYTNIMRVLSVS
potCy D----SESTEVILEAALKSAVTRGLCKPGDAVV--ALHRIGSASVIKICVVK
Anid NVKIWQEDVDRRLKWGINHGLKLGIINKGDNIVCVQGWRGGMGHTNTVRVVPAEENLGLSEE
Anig NVKVWQEDVDRRLKWGINHALKLGIINKGDNIVCVQGWRGGMGHTNTVRVVPAEENLGLAEE
yea D---WTDDVEARINFGIEKAKEFGILKKGDTYVSIQGFKAGAGHSNTLQVSTV
TbrCy1 D-----KDKEKRVKLGLDFAKKEKYASTGDVVVVVHADHSVKGYPNQTRLIYLP
TbrCy2 D-----KDKEKRVKLGLDFAKKEKYASTGDVVVVVHADHSVKGYPNQTRLIYLP
Eco ------FSTDDFYRLGKELALQSGLAHKGDVVVY-GFWCTGTERHY
Bst ------NTTDEMLDVAVDAAVRSGLVKHGDLVVITAGVPVGETGSTNLMKVHVISDLLAKGQ
Bst GIGASRRSARPL

leads to a decrease of activity which can be reversed by dephosphorylation. The other types of pyruvate kinase have different N-terminal sequences and are thus not susceptible to control by phosphorylation.

Some eubacterial pyruvate kinases are also allosterically regulated. In *E. coli* there are two separate pyruvate kinases, both of which are subject to allosteric control. One form is activated by fructose 1,6-*bis*phosphate whilst the other is activated by AMP, ribose 5-phosphate and other nucleoside monophosphates (Speranza *et al.*, 1989). *Bacillus stearothermophilus*, a moderately thermophilic eubacterium, has only one pyruvate kinase, that which is regulated by AMP (Sakai *et al.*, 1986).

It has recently been demonstrated that the allosterically regulated mammalian M2 isoenzyme is encoded by the same structural gene as the M1 isoenzyme which displays hyperbolic kinetics under physiological conditions (Noguchi *et al.*, 1986). The gene contains 12 exons two of which, exons 9 and 10, are differentially spliced (by 'exon-swapping') in a tissue-specific manner. These two exons encode a stretch of 54 amino acids of which 33 remain identical in each exon. The 54 residues form two helices, designated Cα1 and Cα2, which reside in the C domain of the enzyme. X-ray crystallographic studies have shown that these two helices come together to form the 1,3 intersubunit interface of the protein. Since the two isoenzymes differ only in this region of intersubunit contacts, it has been speculated that it is this region of the protein which confers allosteric properties to the M2 isoenzyme. Similarly, the L and R isoenzymes are also identical except that the N-terminal region of the R isoenzyme is longer. In this case, the use of alternative promoters yields isoenzymes of different lengths (Noguchi *et al.*, 1987).

1.4.3 The structure of pyruvate kinase.

Sequence information for all four of the mammalian isoenzymes is available as are pyruvate kinase sequences from a phylogenetically wide range of organisms (Table 1.5, to date, no complete sequence is available

source	nature	size	reference
Cat M1	Protein	530	Muirhead *et al.* (1986)
Chick M1	cDNA	529	Lonberg and Gilbert (1983)
Rat M1	cDNA	530	Noguchi *et al.* (1985)
Rat M2	cDNA	530	Noguchi *et al.* (1985)
Human L	cDNA	543	Tani *et al.* (1988)
Rat L	cDNA	543	Lone *et al.* (1985)
Rat RBC	cDNA	574	Noguchi *et al.* (1987)
Yeast	gDNA	500	McNally *et al.* (1989)
A. nidulans	gDNA	526	de Graaf and Visser (1988)
A. niger	gDNA	526	T. Murcott (personal comm.)
Trypanosome	gDNA	526	T. Murcott (personal comm.)
B. stearo-thermophilus	cDNA	582	Sakai *et al.* (1986)

Table 1.5: Pyruvate kinase complete sequences currently held in the database.

for an archaebacterial pyruvate kinase). Pairwise comparisons of the sequences aligned in Fig. 1.7 (expressed as percentage amino acid differences) are shown in Matrix 1.1.

Three dimensional structural information for pyruvate kinase has been made available by X-ray crystallographic studies of the M1 isoenzyme isolated from cat muscle. The structure has been resolved to 2.6 Å (Muirhead *et al.*, 1986) and is shown in Fig. 1.8. Fig. 1.8 shows a monomer of the enzyme whilst Fig. 1.9 shows the tetrameric arrangment of the monomers in the native enzyme. Each monomer is composed of four domains, N, A, B and C, which are arranged as shown in Fig. 1.9. Domain N consists of the first 42 N-terminal residues, the first nine of which cannot be assigned in the electron density map - a reflection of their structural flexibility. Domain A (residues 43-115 and 224-387) consists of an eight-stranded α/β barrel structure, a motif which has been found in 16 other enzymes including the glycolytic enzymes triose phosphate isomerase, fructose 1,6-*bis*phosphate aldolase and enolase (it is interesting to note that, given the small number of protein structures elucidated to date, α/β barrel enzymes account for 10% of known enzymes). The structure consists of eight β strands alternating with one or two α helices. The chain folds such that the eight strands form a parallel β sheet which twists to give a closed barrel structure and the α helices then pack onto the exterior of the barrel. In pyruvate kinase, unlike other α/β barrel enzymes, the barrel is not contiguous - the main chain is interrupted, folding into domain B, before returning to complete the barrel. The evolution of the α/β barrel motif is a moot subject, the controversy surrounding it being based on a convergent versus divergent evolution argument (Chothia, 1988). Many workers believe that all 17 enzymes may have diverged from a common ancestor, despite their lack of functional and sequence similarities (Farber and Petsko, 1990). However, the recent discovery of an α/β protein which has no enzymic activity - the seed storage protein narbonin from *Vicia narbonensis* (Hennig *et al.*, 1992) - lends support to the idea that the 8-stranded barrel conformation is a stable building block onto which a

	hm2	rm2	cm1	chm	hml	rtl	rtr	pot	And	Ang	yea	Tb1	Tb2	Eco	Bst
hummus2	100	97	93	86	71	70	70	44	54	54	51	51	51	49	47
ratmus2		100	92	86	70	69	69	44	54	54	51	51	51	49	47
catmus1			100	88	69	67	67	43	52	52	49	50	50	49	47
chimus				100	69	69	69	43	52	52	49	51	51	48	48
humliv					100	92	92	44	51	51	50	49	49	47	46
ratliv						100	99	42	51	50	50	47	47	46	45
ratrbc							100	42	51	50	50	47	47	46	45
pot								100	43	43	39	41	41	47	45
Anid									100	94	67	47	47	45	45
Anig										100	66	47	48	46	46
yea											100	48	48	44	42
TbrCy1												100	99	44	41
TbrCy2													100	44	41
Eco														100	51
Bst															100

Matrix 1.1: Pairwise comparisons of pyruvate kinase sequences. Abbreviations: *hummus* - human muscle, *ratmus* - rat muscle, *catmus* - cat muscle, *chimus* - chicken muscle, *humliv* - human liver, *ratliv* - rat liver, *ratrbc* - rat erythrocyte, *pot* - potato, *Anid* - *Aspergillus nidulans*, *Anig* - *Aspergillus niger*, *yea* - *Saccharomyces cerevisiae*, *TbrCy* - Trypanosome cytoplasm, *Eco* - *E. coli*, *Bst* - *Bacillus stearothermophilus*.

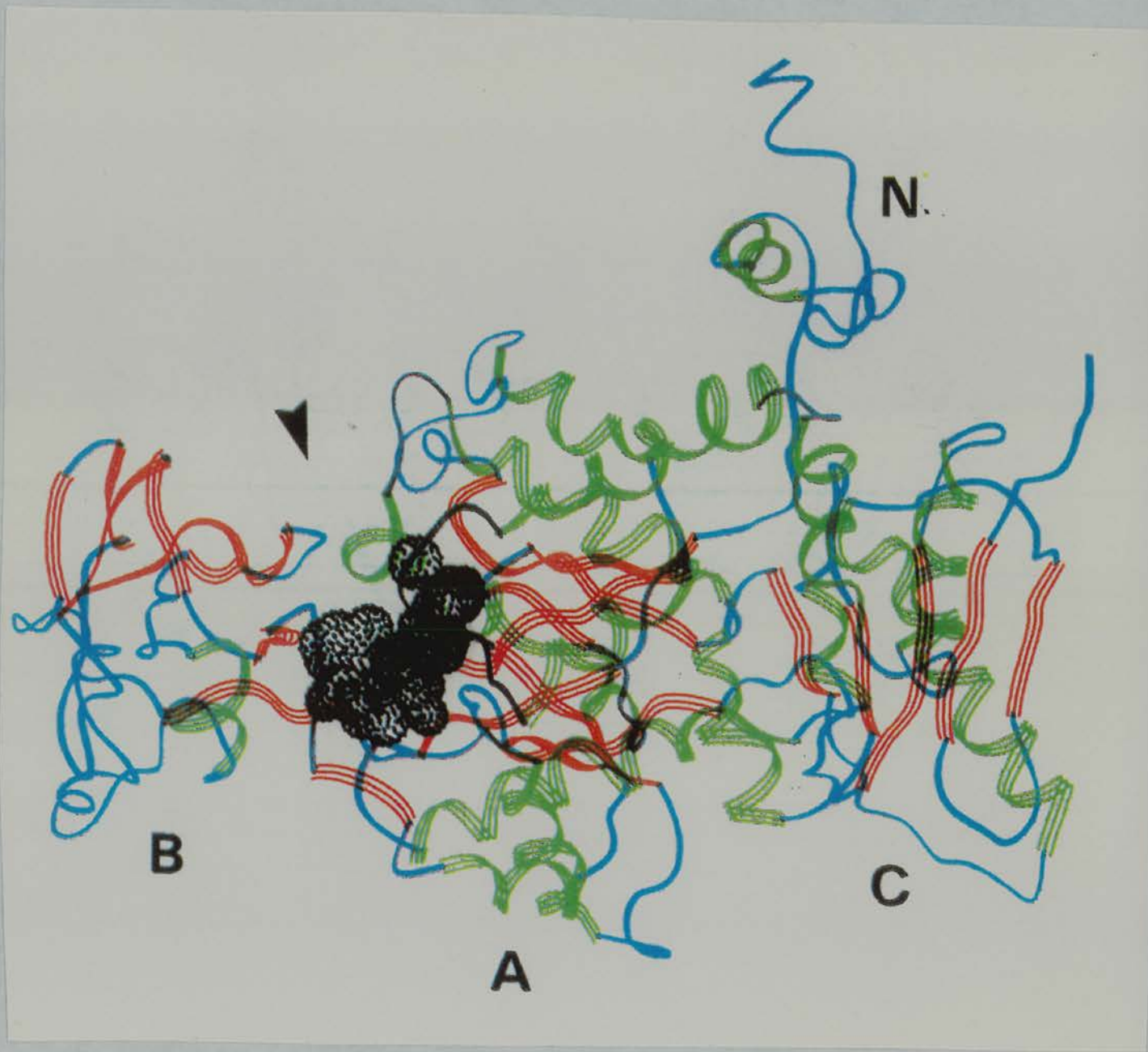


Figure 1.8: The subunit of cat muscle pyruvate kinase. The four domains (N, A, B, and C) are labelled, and ligand access to the active site is indicated. The likely location for the effector site is at the base of the N domain. Space filling models of ATP and pyruvate are shown at the active site. The coordinates were made available by H. Muirhead. Red: β-sheet; green: α-helix; blue: coil.

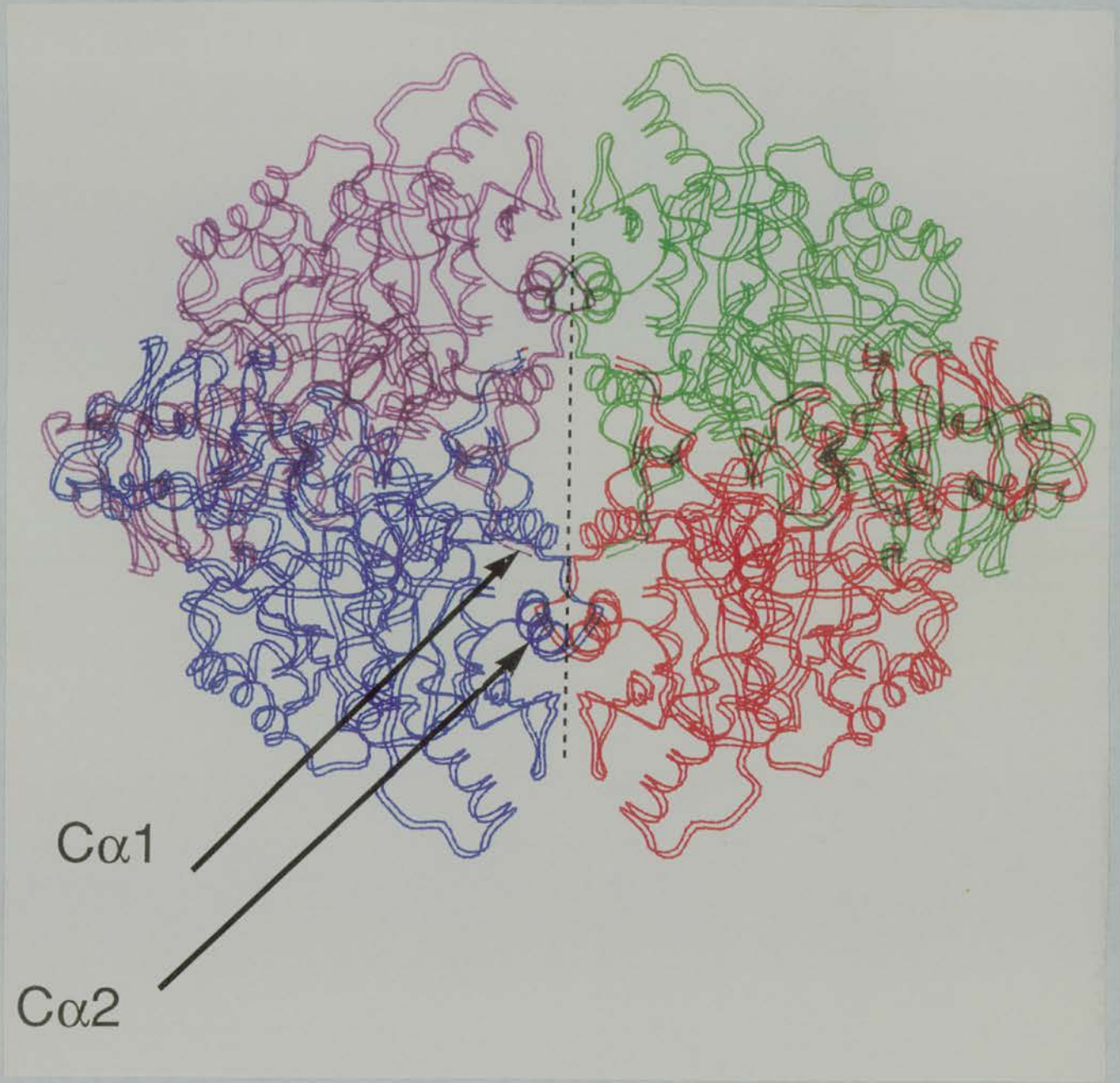


Figure 1.9: Tetramer of cat muscle pyruvate kinase. The broken line shows the 1,3-intersubunit interface. The helices marked Cα1 and Cα2 are thought to be involved in conferring allosteric properties on the enzyme.

number of independently evolved functions have converged. This latter theory does seem particularly likely given the diverse nature of the proteins which possess this fold and the fact that the fold in enolase is formed (from the primary structure) rather differently to that of the other barrel proteins.

The active site of the pyruvate kinase lies in a cleft between domains A and B (residues 116-223) and the allosteric effector sites are probably (there is no direct evidence for this) between domains A and C (residues 388-530). The tetramer is put together in such a manner that it is effectively a dimer of dimers, there being two different types of intersubunit contact within the native molecule. The 1,2 intersubunit contact involves both domains A and C whereas the 1,3 contact region involves only domain C. The effects of altering the residues within these contact regions on both the quaternary structure and allosteric properties continue to be the subject of intensive study (McNally and Fothergill-Gilmore, 1990; Collins *et al.*, 1992).

1.4.4 The evolution of pyruvate kinase

Pyruvate kinase is a strongly conserved enzyme. The diagram in Fig. 1.10 shows the evolution rate for the enzyme which averages to about 8 PAMs per 100 million years. This is an intermediate rate for a glycolytic enzyme: enolase and glyceraldehyde 3-phosphate dehydrogenase evolve slowest (5 PAMs/100 million years) whilst the N-terminal half of hexokinase evolves the fastest (16 PAMs/100 million years). The curve of Fig. 1.10 may be biphasic in nature - a puzzling observation which is discussed in the general discussion section at the end of this thesis. If pyruvate kinase is one of the progenitor enzymes of the pathway, how might it have further evolved into the other enzymes of glycolysis? Indeed, by what processes did pyruvate kinase evolve to become the enzyme which we recognise today? These questions can be answered by considering the basic principles of protein (and catalyst) evolution. This process can be considered to be occurring at various levels from single amino acid changes to extensive alterations involving the entire molecule.

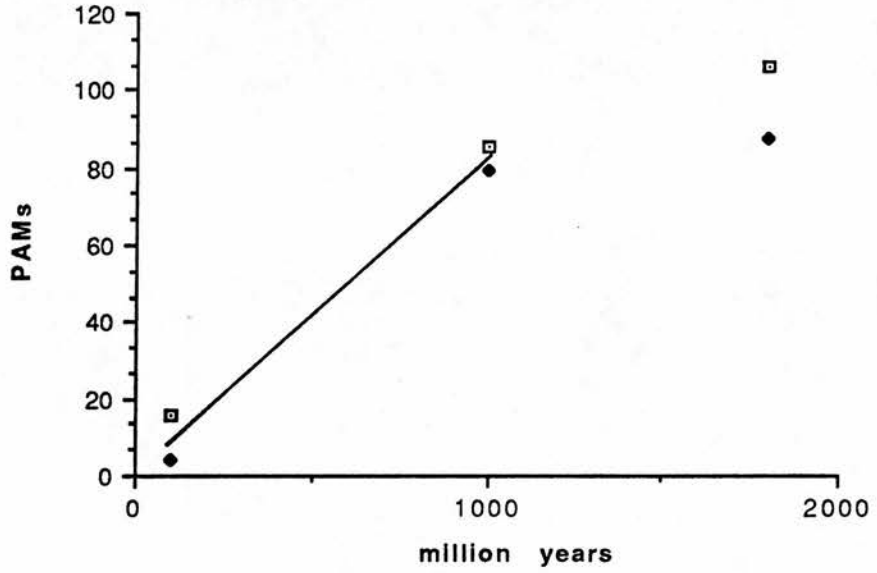


Figure 1.10: Evolution rate of pyruvate kinase. The graph shows the differences between amino acid sequences of the enzyme (as PAMS) against organism divergence times for the enzyme from selected organisms. The figure is discussed in Chapter 8. Reproduced from Fothergill-Gilmore and Michels (1992).

The DNA of any organism is constantly subjected to the influence of a variety of mutagenic agents from chemicals to ultra violet radiation. The chemical changes incurred in the DNA by the action of such agents are often deleterious to the organism and do not persist (i.e. they do not become 'fixed'). Others, however, are neutral in their effects (or, more rarely, advantageous) and may be seen phenotypically as amino acid changes in proteins. Such changes seem to occur at a steady rate (this led to the molecular clock hypothesis of Wilson, 1985). Some amino acids change relatively frequently, the most frequent being a glutamate to aspartate change, whilst other changes seem to be 'forbidden'. It is worth noting that glycine residues are not often changed to something else. However, if one amino acid is changed to another, the highest probability is that it will mutate to a glycine. This may be due to the small size of the glycine moiety. Point mutations of this type can explain many of the changes noted between pyruvate kinase sequences. They may even explain why in two cases, mentioned earlier, the enzyme exists as a dimer rather than as a tetramer - if the residues in the second subunit interface region were significantly altered, the tetramer could fall apart into two dimers.

Point mutations, however, cannot explain some more drastic differences between pyruvate kinases from various organisms. The enzyme from rat erythrocytes is 48 residues longer than that from trypanosomes. This may be due either to an insertion (or deletion) event at some point after the separation of the ancestors of the two organisms or to the use of alternative promoters in the L/R gene. Almost without exception such events occur in surface loops where they can be tolerated without disastrous disruption of structure or loss of function. Another more prominent change is in the region of rat M1 and M2 isoenzymes between residues 380 and 435 (see 1.4.2). This change has resulted due to an exon duplication followed by differential splicing. Over time the two exons have diverged in sequence until they are only about 50% identical to one another.

The jigsaw describing the evolution of pyruvate kinase, whilst it definitely includes the processes described above, is missing a piece. No

complete sequences are yet available for the enzyme from archaebacterial sources. Information of this kind could be extremely illuminating especially in the light of evidence gained from studies of other archaebacterial glycolytic enzymes. Archaebacterial phosphoglycerate kinases have been demonstrated to display 30-36% identity to their bacterial and eukaryotic counterparts (Fabry *et al.*, 1990). This degree of similarity is consistent with the enzymes from all domains sharing a common ancestor. In contrast, the sequences of archaebacterial glyceraldehyde-3-phosphate dehydrogenases exhibit only 7-15% identity with their counterparts from the other domains (Hensel *et al.*, 1989). This could suggest that the enzyme from archaebacteria does not share a common ancestor with those isolated from bacterial and eukaryotic sources but is, in fact, a different protein which acts on the same substrates/products - i.e. it is an example of convergent evolution. One of the main aims of this project is to obtain sequence information from an archaebacterial pyruvate kinase in order to help clarify the evolutionary path taken by the enzyme.

1.5 *Thermoplasma acidophilum*

1.5.1 Classification of *T. acidophilum*

Thermoplasma acidophilum, a thermophilic archaebacterium, was first isolated from a burning coal-refuse pile in Indiana (Darland *et al.*, 1970). Due to the lack of a cell wall the organism was originally thought to be a mycoplasma (hence the name) but was subsequently recognised by 16S rRNA sequence comparisons to cluster with the archaebacteria. It has since been discovered in hot springs and sulphataric fields (Segerer *et al.*, 1988) thriving at temperatures between 50-64^{o}C and pH's between 0.5 - 3.0. Two members of the genus *Thermoplasma* are currently recognised: *T. acidophilum* and *T. volcanium*.

T. acidophilum, phylogenetically classified on the basis of its rRNA, belongs to the methanogen-halophile branch (euryarchaeota) of the archaebacterial tree, despite its phenotypic resemblance to the chrenarchaeota. Contrasting with this deduction, analyses of the RNA

polymerase subunit structure (Prangishville *et al.*, 1982) and of the functional properties of the ribosome (Amils *et al.*, 1989) suggest that *T. acidophilum* more closely resembles the sulphur-dependent thermophiles of the chrenarchaeota than representatives of the euryarchaeota. Further to these findings, *T. acidophilum* is unique amongst the archaebacteria in possessing unlinked rRNA genes (Tu and Zillig, 1982) - other archaebacteria have linked rRNA genes, typically in the order 5'16S-23S-5S 3' (Neumann *et al.*, 1983). The genus *Thermoplasma*, therefore, presents a taxonomic problem. It is entirely possible that the organism represents a link between the two main branches of the archaebacterial tree and may even constitute a third branch of its own.

1.5.2 *T. acidophilum* and its environment.

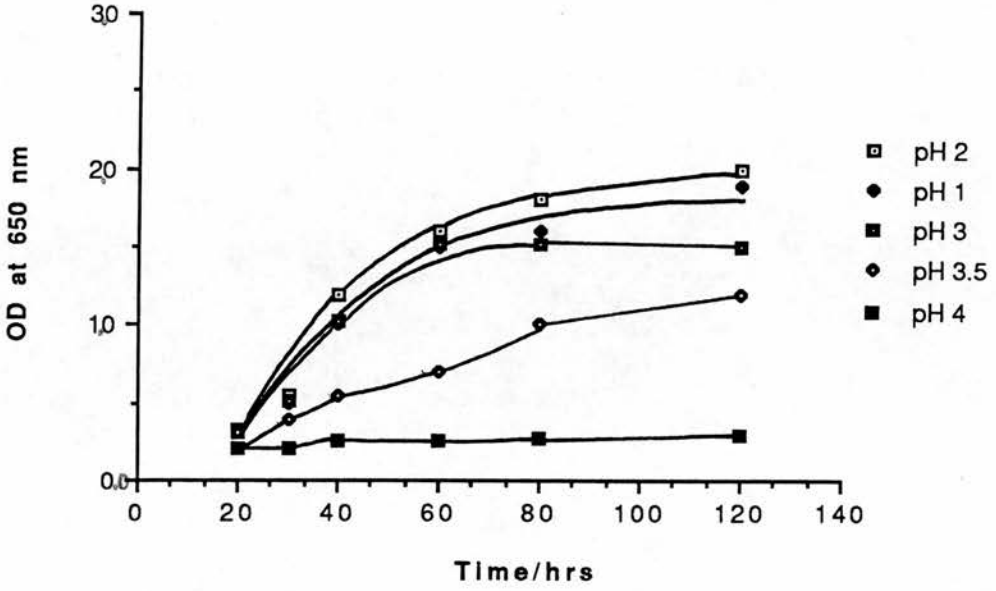
T. acidophilum grows between pH 0.5-3.0 and at temperatures of between 37-65^{o}C. Figs. 1.11 and 1.12 show the dependence of the organism's growth rate upon both parameters. Fig. 1.11 shows a plot of the growth rate in doublings per hour against the reciprocal of absolute temperature. The optimum growth temperature is about 59^{o}C. No growth occurs at 65^{o}C or 37^{o}C. The effect of pH on the growth of the organism is plotted in Fig. 1.12. No growth occurred at pH 0.35 and growth was very slow at pH 4. The organism does grow well, however, over the range in between with an optimum of pH 1-2.

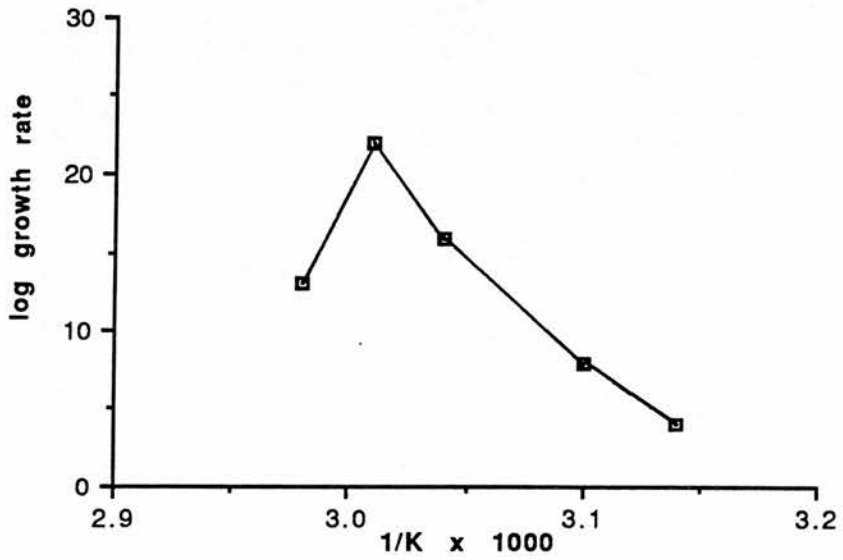
Despite the acidic conditions of its growth environment, *T. acidophilum* maintains an internal pH of around 6.0. Other archaebacterial acidophiles regulate their internal pH by the use of an H^{+}-translocating ATPase. This enzyme is missing from *T. acidophilum*, however, which instead employs an electron transport chain (Searcy and Whatley, 1982). Searcy has further suggested that this might be because the environmental conditions in which *T. acidophilum* exists may be too extreme for an ATPase (Searcy, 1986).

The membrane lipids of *T. acidophilum* are typical of archaebacterial lipids: ether-linked C40 isoprenyl glycerolipids (Langworthy *et al.*, 1972).

Figure 1.11: Effect of pH on the growth of *T. acidophilum.* The organism was grown at 55^{o}C in the standard growth medium. Growth was followed by measuring the optical density at 650 nm in a Spectronic 20. Taken from Darland *et al.* (1970).

Figure 1.12: Arrhenius plot of the log of the growth rate vs the reciprocal of the absolute temperature. Cultures were grown without aeration in the standard growth medium at pH 2.0. Growth was followed by measuring the optical density at 650 nm. Taken from Darland *et al.* (1970).


OD at 650 nm
30
20
10
0.0
0
20
40
60
80
100
120
140
Time/hrs
pH 2
pH 1
pH 3
pH 3.5
pH 4


log growth rate
30
20
10
0
2.9
3.0
3.1
3.2
1/K x 1000

Their unusual, cross-linked structure enables the formation of lipid monolayers rather than the more common bilayers and may, therefore, be an adaptation which stabilises the plasma membranes at elevated temperatures. The use of ether rather than ester linkages enhances the stability of the lipids towards acid-base hydrolysis.

T. acidophilum DNA has an average G-C base composition of 46% (Christiansen *et al.*, 1975) and so it is unlikely to be stabilised at high temperature by having a prevalence of G-C linkages. Instead, the DNA is thought to be stabilised against thermal denaturation under the auspices of a DNA-binding histone-like protein. This protein (HTa) has been purified from *T. acidophilum* and has been found to be similar to eukaryotic histones in terms of its basicity and amino acid composition (DeLange *et al.*, 1981a).

1.5.3 Proteins of *T. acidophilum*

Amino acid sequences are now available for four *T. acidophilum* proteins: the HTa protein discussed in 1.5.2 (DeLange *et al.*, 1981b); ferredoxin (Wakabayashi *et al.*, 1983); glucose dehydrogenase (Bright *et al.*, 1993) and the citrate synthase (Sutherland *et al.*, 1991). Simple sequence comparisons of these proteins and their counterparts from evolutionarily divergent organisms have failed to elucidate any structural features which appear to confer thermostability on the archaebacterial versions. This is not entirely surprising, however, as it has been shown that the addition of a single hydrogen bond or salt bridge on the surface of a protein can result in a significant increase in the protein's thermostability (Danson, 1988). More informative are three-dimensional structure comparisons, especially when coupled with site-directed mutagenesis experiments. Such studies have yielded some useful results which will be discussed later.

1.5.4 *T. acidophilum* as a 'protoeukaryote'

It has been suggested that a thermoacidophilic archaebacterium in general, and *T. acidophilum* in particular, may represent the closest

living counterpart of the ancestor of all eukaryotic cells (Searcy, 1986). This observation, prompted by the closeness in most phylogenetic trees of the crenarchaeota (or eocytes) to the eukaryotic branches, may be supported by several lines of evidence: there are three particular features of *T. acidophilum* which point to an evolutionary path of this type.

First, the archaebacterium lacks a H^+-translocating ATPase. This adaptation to high environmental temperature and low pH has led to a loss of oxidative phosphorylation and excretion of acetic acid by the organism. This may have been a preadaptation (Searcy's term - presumably implying a chance event which was a prerequisite for another event) to the acquisition of a symbiotic bacterium, initially for the disposal of acetic acid. Secondly, the archaebacterium lacks a rigid cell-wall. This phenomenon, possibly an adaptation to lithotrophic metabolism requiring membrane contact with mineral substrates, led to the development of a Ca^{2+}-regulated cytoskeleton based upon actin and myosin. This might have been a preadaptation to amoeboid cellular movement, cytoplasmic streaming, endocytosis and exocytosis. Thirdly, *T. acidophilum* contains eukaryotic- like DNA-binding proteins. This adaptation to enhance the thermal stability of the DNA genome has led to the condensation of the DNA into nucleosome-like particles. This may have been a step on the way to the development of the condensed chromosomes seen in eukaryotic cells today.

It is worth noting, as a *caveat* to this speculation, that the thermoacidophilic archaebacteria have continued to adapt to their harsh environments throughout evolution and are, therefore, likely to be more reduced and specialised than were their ancestors. Some workers believe that such adaptations occurred entirely since the separation of the lineages of the three domains and that thermophily is a relatively recent innovation (Forterre, 1992). Nevertheless, it is possible that a thermophilic archaebacterium, perhaps in conjunction with an ancestral bacterium, may have been the ancestral eukaryote.

1.6 Aims of the project

The central aim of this project was to improve the understanding of the molecular processes behind the evolution of the enzymes of glycolysis and of the glycolytic pathway as a whole. It was decided to gain information about the pyruvate kinase of the thermoacidophilic archaebacterium *Thermoplasma acidophilum* in order to compare the physical, kinetic and structural properties of the enzyme from that organism with counterparts isolated from bacterial and eukaryotic sources.

Pyruvate kinase was chosen for two reasons. First, there is a wealth of sequence and structural information available for the enzyme isolated from many different sources. Secondly, it is thought to be one of only two glycolytic enzymes which are ubiquitous and, therefore, potentially useful as phylogenetic marker molecules. *Thermoplasma acidophilum* was chosen because it was known to possess a pyruvate kinase activity in the modified Entner-Doudoroff pathway of glycolysis and because it occupies a very interesting position in recent phylogenetic trees.

1.7 Strategy of the project

Thermoplasma acidophilum was batch cultured in order to provide enough cells for the purification of the pyruvate kinase. The enzyme was then purified using a multi-step chromatographic procedure and physically and kinetically characterised. Of particular interest were the quaternary organisation of the protein (which was assessed by native polyacrylamide gel electrophoresis and gel filtration chromatography) , its thermal stability (assessed by assaying the enzyme after incubations at various temperatures) and its control mechanisms (i.e. the dependence of the activity of the enzyme on the concentrations of its substrates was determined as were the effects of known allosteric effectors of other pyruvate kinases).

The gene for the enzyme was then cloned and partially sequenced by

dideoxy (Sanger) DNA sequencing for alignment with other pyruvate kinase sequences and subsequent phylogenetic analysis using the program CLUSTAL. The subsequent chapters of this thesis describe the methods used to fulfil these ends, the rationale behind the methods chosen, and the results obtained at each stage.

Chapter 2: Materials and Methods

CHAPTER 2: MATERIALS AND METHODS

2.1 Materials

2.1.1 Strains

Thermoplasma acidophilum	DSM 1728
Escherichia Coli	TG1: K12, D(lac-pro), supE, thi, hsdDS/F'traD36, proA+B+, lacIq, lacZDM15

2.1.2 Vectors

vector	source
pBluescriptII M13 SK+	Stratagene, La Jolla, CA
pCRII	Invitrogen, San Diego, CA

2.1.3 Growth media

(a) Oxoid Ltd., Haverhill, Suffolk
Agar No.1, yeast extract

(b) Sigma Chemical Company, Poole, Dorset
Kanamycin, Xgal, IPTG, all trace components

2.1.4 Radiochemicals

Amersham p.l.c., Aylesbury, Bucks.

Deoxyadenosine 5'-(α-^{33}P)triphosphate: 370 MBq/ml, 10 mCi/ml; specific activity 3000 Ci/mmol, 37-110 TBq/mmol

2.1.5 Enzyme assays and purification

(a) Boehringer Mannheim, Lewes, E. Sussex
pyruvate kinase, lactate dehydrogenase, 3,4-dichlorisocoumarin

(b) Sigma Chemical Company, Poole, Dorset
Isocitrate substrate, PEP, ADP, ATP, Imidazole, 5'-AMP agarose, E64c

(c) Pharmacia, Milton Keynes
Superose 6, Mono-Q

(d) Aldrich Chemical Company, Gillingham, Dorset
1,10-phenanthroline

2.1.6 Enzymes

(a) Amersham p.l.c., Aylesbury, Bucks.
EcoRI, HinDIII, BamH1, terminal transferase

(b) Boehringer Mannheim, Lewes, E. Sussex
T4 DNA ligase, RNaseA, Proteinase K, clostripain

(c) Promega Ltd., Southampton
Taq polymerase

2.1.7 Oligonucleotides

All oligonucleotides were synthesised by the OSWEL DNA service of the University of Edinburgh

2.1.8 Miscellaneous

(a) Amersham p.l.c., Aylesbury, Bucks.
Hybond-N roll, Hyperfilm MP X-ray film, ECL-reagents and system

(b) Promega Ltd., Southampton
fmol sequencing kit

(c) Invitrogen, San Diego, CA
TA cloning kit

(d) Sigma Chemical Company, Poole, Dorset
Polaroid 665 film, agarose

All other biochemicals mentioned were supplied by Sigma or BDH Ltd., Poole, Dorset.

2.2 Methods

2.2.1 Growth and harvesting of *Thermoplasma acidophilum*

Cells were supplied by the German culture collection (Deutsche Sammlung von Mikroorganismen, DSM) lyophilised in a glass vial. They were resuspended in 2 ml of Growth Medium (as recommended by DSM's 1989 catalogue of strains): $(NH_4)_2SO_4$, 1.32 g; KH_2PO_4, 0.372 g; $MgSO_4 X 7H_2O$, 0.247 g; $CaCl_2 X 2H_2O$, 0.074 g; Trace element solution, 10.0 ml (see below); Yeast extract, 1.0 g; Glucose, 10.0 g; Distilled water,1000.0 ml. The pH was adjusted to 2.4 with 10 N hydrochloric acid. Stock solutions of yeast extract (10% w/v) and glucose (50% w/v) in water were sterilized separately by autoclaving and then added to the sterile mineral salt

medium.

The trace element solution was composed of: $FeCl_3 X 6H_2O$, 1.93 g; $MnCl_2 X 4H_2O$, 0.18 g; $Na_2B_4O_7 X 10H_2O$, 0.45 g; $ZnSO_4 X 7H_2O$, 22.0 mg; $CuCl_2 X 2H_2O$, 5.0 mg; $Na_2MoO_4 X 2H_2O$, 3.0 mg; $VOSO_4 X 5H_2O$, 3.8 mg; $CoSO_4 X 7H_2O$, 2.0 mg; Distilled water, 1000.0 ml.

The 2 ml cell suspension was then used to seed a 20 ml starter culture incubated at the growth temperature of 59°C in a 50 ml culture flask placed on its side. This was allowed to grow for four days or until the culture had reached an A_{650} of 0.20. 1.5 ml of this culture was then used to inoculate a 2 l flask of growth medium. After 50hrs of incubation at 59°C the cells were ready for harvesting.

Cells were harvested by centrifugation in 250ml bottles at 3250 g (5000 r.p.m. in a JA14 rotor) for 20 min. Cell pellets were resuspended in growth medium and lysed (see 2.2.3) for immediate use or frozen in 1ml aliquots for later use.

2.2.2 Preparation of cells for electron microscopy

Cells were prepared for electron microscopy using a modification of the method of Darland *et al.* (1970). Cells were fixed in 1ml 0.1 M cacodylate buffer (pH 6.1) containing 6% (v/v) glutaraldehyde for 4 hr at 4°C and then centrifuged at 9150 g (13000 r.p.m. in a bench microfuge). The pellet was resuspended and washed in 0.1 M cacodylate buffer containing 0.25 M sucrose and cells were then transferred to 2% (w/v) osmium tetroxide in Ryter-Kellenberger buffer (see below). After fixation in this solution for 18 hr at 4°C, the cells were resuspended in 0.04 ml of 2% (w/v) agar and washed for 2 hr in 0.5% (w/v) uranyl acetate in Ryter-Kellenberger buffer. This was followed by dehydration with graded ethanol (50 - 100% v/v) and passage of the cells through propylene oxide. Cells were embedded in araldyte and baked at 60°C overnight to solidify the mix. The block was then sectioned using a microtome and stained with lead citrate.

Composition of Ryter-Kellenberger buffer:
STOCK - sodium barbitone, 2.94 g; sodium acetate, 1.94 g; sodium

chloride, 3.40 g, made up to 100 ml with distilled water. The buffer used was 5 ml stock + 13 ml distilled water with $CaCl_2$ added to a concentration of 1.0 M.

2.2.3 Preparation of cell-free extract

200 mg of fresh cells were resuspended in 2 ml of lysis buffer (NaCl, 0.02 M; $MgCl_2$, 0.01 M; Tris HCl, 0.01 M). The following protease inhibitors were added prior to cell disruption in order to prevent degradation of the enzyme:

i) 1,10-phenanthroline - inhibits metalloenzymes
Stock solution: 100 mM in DMSO, diluted x 1000 for use

ii) E64C - inhibits cysteine proteases
Stock solution: 1 mM in SDW, diluted x 50 for use

iii) 3,4-Dichlorisocoumarin - inhibits serine proteases
Stock solution: 5 mM in DMSO, diluted x 50 for use

The final pH of the buffer was adjusted to 7.7. The resuspended cells were then homogenised at room temperature using glass beads. The homogenate was centrifuged at 9150 g (13000 r.p.m. in a bench microfuge) for 10 min. The pellet was discarded and the supernatant assayed for pyruvate kinase activity.

2.2.4 Purification of *Thermoplasma* pyruvate kinase

The purification of the pyruvate kinase involved a novel multi-step procedure incorporating ammonium sulphate precipitation followed by ion-exchange, gel filtration and affinity chromatography. Where possible, all steps were performed at 60°C. All columns were jacketed to this temperature.

a) Ammonium sulphate precipitation: The supernatant from cell breakage was prepared as in section 2.2.3. Ammonium sulphate was then slowly added to this supernatant up to 35% saturation and left stirring at room temperature for 20 minutes. The resultant suspension was centrifuged at 10900 g (9500 r. p. m.) in a JA20 rotor for 30 minutes and the pellet discarded. The supernatant of this spin was then made up to

65% saturation with ammonium sulphate. After stirring at room temperature for 20 minutes this suspension was also spun under the same conditions as above. The pellet of this spin, containing the pyruvate kinase, was saved and the supernatant discarded. The pellet was then redissolved in 1 ml of 10 mM Tris buffer (pH 7.5) containing 20 mM NaCl and 10 mM $MgCl_2$ and the solution then dialysed overnight at room temperature against 1l of the same buffer.

b) Ion-exchange chromatography: Two buffers were used for the ion-exchange column: Buffer A (NaCl, 20 mM; $MgCl_2$, 10 mM; Tris, 10 mM; pH 7.5) and buffer B (NaCl, 300 mM; $MgCl_2$, 10 mM; Tris, 10 mM; pH 7.5). Both buffers were filtered under pressure through a 0.2 μm filter and stored at 4^oC until required (note that buffer A is equivalent to lysis buffer without the protease inhibitors).

The dialysate from step (a), already suspended in buffer A, was added to the top of a Mono-Q column (7 mm x 42 mm), attached to the Pharmacia f. p. l. c. system. A pre-programmed gradient was then used to elute the column. Parameters of the gradient were as follows: flow rate 1 ml/minute; 100% A (0% B) for 5 minutes; 0 - 100% B (linear gradient) over 30 minutes; 100% B (0% A) for 10 minutes. Elution of the proteins in the sample was followed by a UV monitor set at 280 nm. Fractions of 1 ml were collected and those containing pyruvate kinase activity pooled for further purification.

c) Gel-filtration chromatography: The pooled pyruvate kinase fractions from the Mono-Q column were applied to a Superose 6 column (15 mm x 285 mm) attached again to the f. p. l. c. system. The column was then eluted isocratically with buffer A at a flow rate of 0.5 ml/minute (this procedure served also to desalt the sample - column elution followed at 280 nm). Fractions of 1 ml were collected and those containing pyruvate kinase activity pooled for application to the affinity column. Both f. p. l. c. columns used were prewashed with the appropriate buffer as recommended by the manufacturers. After use they were again washed for 1 hr and then stored either in 80% methanol or ethanol.

d) Affinity chromatography: A small (4 ml volume: 10 mm x 50 mm) column of N^6 (6-aminohexyl-)5' AMP-sepharose 4B was poured after being degassed at 45 ^{o}C and allowed to swell (5 g swells to 20 ml in 0.1 M phosphate buffer, pH 7.0). The column was then prewashed with five volumes of the phosphate buffer and then the enzyme pool from the Superose 6 column added to the top of the column. Unbound material was washed from the column with four volumes of the phosphate buffer and then the buffer was changed to one containing 10 mM ADP. Two volumes of this buffer was adequate to elute all pyruvate kinase activity from the column. The eluted enzyme was concentrated in an Amicon concentrator and an aliquot applied to SDS-PAGE gel electrophoresis. The enzyme was found to be homogeneous and free of any contaminants. This preparation was then subjected to further characterisation and sequencing studies.

2.2.5 Stabilisation of *Thermoplasma* pyruvate kinase

The instability of the purified pyruvate kinase of *Thermoplasma* was a problem which became more critical as the storage temperature was lowered. Since storage at 40 ^{o}C or above was considered to be impracticable, a method of stabilising the enzyme at room temperature was required. In 1976, Yun *et al.* developed a preservation buffer for pyruvate kinases and this buffer was adapted to be suitable for preserving the archaeal enzyme. The final elution buffer from the affinity column had the components below added to the following concentrations to make up the stabilisation buffer: 0.1 M sodium phosphate, 25% glycerol, 5 mM EDTA, 5 mM β-mercaptoethanol.

2.2.6. Assays for pyruvate kinase

Two assays for the enzyme were used:

2.2.6.1. The coupled assay of Bucher and Pfleiderer (1955)

For a given assay, the quartz cuvette contained in 1 ml: $MgSO_4$, 15 mM; KCl, 100 mM; Imidazole, 50 mM; 0.1 mg PEP, 0.1 mg ADP, 0.1 mg NADH and 0.1 mg lactate dehydrogenase. The pH of the assay buffer was adjusted to 7.6 and the reaction started by addition of 10 µl of the pyruvate kinase

solution. The conversion of NADH to NAD^+ was followed at 340 nm in a Phillips recording spectrophotometer.

2.2.6.2. The direct assay of Pon and Bondar (1967)

For an assay the cuvette contained in 1 ml: $MgSO_4$, 7.2 mM; KCl, 72 mM; Tris, 50 mM; ADP and PEP (added to varying concentrations for kinetic studies). In any one series of assays, the components with invariant concentrations were mixed as a stock solution, prepared freshly each day, and kept on ice. The reaction was started with the addition of 10 µl of the pyruvate kinase solution and the disappearance of PEP followed at 230 nm in a Phillips recording spectrophotometer.

2.2.7. Estimation of Protein

Protein was determined by a modification of the method of Bradford (1976). A stock solution was prepared by dissolving 100 mg of Coomassie Brilliant Blue G in 50 ml 95% ethanol and adding 100 ml 85% orthophosphoric acid. 18ml of this stock solution was diluted to 100ml with SDW and then filtered through Whatman No.1 filter paper.

For protein estimation 0.5 ml of the protein sample (containing 10-50 µg of protein) was added to 2.5 ml of the diluted Bradford protein reagent and the absorbance measured at 595 nm after incubating at room temperature for 30 minutes. A standard curve was constructed using bovine serum albumin as the standard.

2.2.8 Assay for isocitrate dehydrogenase

This assay (after Danson and Wood, 1984) is based on following spectrophotometrically the conversion of either NAD^+ or $NADP^+$ to NADH and NADPH respectively at 340 nm. The reaction took place in a 1 ml quartz cuvette containing 0.6 ml assay buffer (Tris-HCl, 20 mM; EDTA, 1.0 mM; $MgCl_2$, 10 mM; $NADP^+$, 0.2 mM or NAD^+, 1.0mM, pH 8.0) and 0.4 ml 4 mM Isocitrate Substrate (Sigma). 10 µl of the isocitrate dehydrogenase solution was added to start the reaction and the increase in A_{340} followed

continuously in a Phillips recording spectrophotometer at 55°C.

2.2.9. Purification of commercial ADP

5'ADP as supplied by Sigma was found to have a significant contamination of AMP such that its concentration in the assay buffer would be 30 μM, enough to partially activate the allosterically controlled pyruvate kinase. The procedure of Hunger and Reinbothe (1974) was used to purify the ADP from the AMP contamination. 1 ml of 10 mg/ml 5'ADP (Sigma) in ammonium carbonate buffer ($NH_4HCO_3^+ + NH_2COONH_4$, 0.07 M) was loaded onto a Pharmacia K9/60 column (0.9 cm x 60 cm) containing 38 ml of DEAE-Sephadex A-25 and the column then eluted with 1 l of a linear gradient from 0.07 M - 0.20 M ammonium carbonate buffer at a flow rate of 42 ml/hr. 10ml fractions were collected and the % transmittance of each determined at 254 nm. Two peaks were detected, a small one at approximately 500 ml containing the AMP and a much larger one around 650 ml containing the purified ADP. Fractions containing the nucleotides were pooled, dried and used in kinetic analyses of the pyruvate kinase.

2.2.10. Determination of the native and subunit Mr of the pyruvate kinase

The Mr of the pyruvate kinase tetramer was determined, according to a method in the Pharmacia booklet 'Gel Filtration Chromatography', using a Sephacryl S-200 gel filtration column (11 mm x 165 mm, matrix volume 62.7 ml) which was equilibrated with buffer A (as used in the purification protocol). The column was calibrated with a series of protein standards of known molecular weights (myosin, 205000; β-galactosidase, 116000; phosphorylase B, 97400; bovine serum albumin, 66000; egg albumin, 45000; carbonic anhydrase, 29000) and the void volume of the column determined with blue dextran (Mr 2,000,000). Elution of the standards was performed isocratically in buffer A using a Pharmacia P-3 peristaltic pump at a flow rate of 1 ml/min; a UV flow cell set at 280nm was used to determine the retention volume of each protein. 1 mg of the affinity-purified pyruvate kinase, in a volume of 1ml, was then loaded onto the column and eluted as before. Fractions of 1 ml volume were collected and each fraction was assayed for pyruvate kinase activity using the coupled assay technique (as described in section 2.2.6).

The Mr of the pyruvate kinase subunits was determined using SDS-polyacrylamide gel electrophoresis. A 12% polyacrylamide gel was poured as described in section 2.2.14. and loaded in lane 1 with 20 µl of high molecular weight standards (Sigma) and in lane 2 with 20 µl of boiled 1 mg/ml affinity-purified pyruvate kinase. The gel was run as described in section 2.2.14. and stained and destained as decribed in section 2.2.15.

2.2.11 Amino Acid Analysis

100 pmol of the purified pyruvate kinase was subjected to automated amino acid analysis using an Applied Biosystems 420A Derivatizer. The protein was acid hydrolysed with 6 N HCl and the amino acids produced reacted with phenyl isothiocyanate. The PTC-amino acid derivatives were then separated on a reverse phase column eluted with a gradient of acetate buffer and acetonitrile. The derivatives were detected at 254 nm. Cysteine residues were detected as pyridylethyl derivatives which were produced by the method below.

2.2.12 Vapour phase pyridylethylation of cysteine residues (Hayes *et al.*, 1989)

100 pmol of the pyruvate kinase (or clostripain peptide) was spotted onto a glass-fibre filter disc and dried under a stream of nitrogen. The pyridylethylation reaction was performed in a stoppered glass tube (50 mm x 10 mm) with a constriction in the middle to support the filter. Pyridine (100 µl), water (100 µl), 4-vinylpyridine (20 µl) and tributylphosphine (20 µl) were placed in the lower part of the tube. The tube was flushed thoroughly with argon and the protein-covered disc was placed on the constriction before the tube was quickly sealed with a stopper. The vapour phase reaction was allowed to continue for 2 hr at 60 ^{o}C. The disc was removed from the tube and washed for 10s in 2ml of each of the following solvents: n-heptane, n-heptane/ethyl acetate (1:1 v/v), ethyl acetate. The washed disc was allowed to dry at room temperature before being wetted with 30 µl (2 mg) of polybrene. The polybrene treated disc was re-dried and placed into the analyser or the sequencer. The pyridylethylcysteine phenylhydantoin derivative is resolved on the PTH C_{18} column with the standard separation protocol used by the on-line 120

analyser.

2.2.13. Amino acid sequencing (Hayes *et al.* 1989)

Peptides from the clostripain digest (see section 2.2.17) were separated and purified using an Applied Biosystems 130A microbore separation system. The column employed was an Aquapore RP-300 (7 mm particle size; 2.1 mm x 30 mm) and it was eluted with a linear 8 - 80% (v/v) gradient of acetonitrile in aqueous 0.1% TFA and monitored at 220 nm. The purified peptides were then subjected to automated sequencing on an Applied Biosystems 477A instrument with a 120A on-line phenylhydantoin analyser. Polybrene (2 mg) was loaded onto a glass fibre filter disc which was then pre-cycled three times before being loaded with approximately 300 pmol of the clostripain cleavage fragment. After Edman degradation (Edman and Begg, 1967), the anilinothiazolinone derivatives cleaved from the peptide were converted automatically into the more stable phenylhydantoin forms and separated on an Applied Biosystems PTH C_{18} (5 mm particle size; 2.1 mm x 220 mm) column that was eluted with a 0-100% (v/v) linear gradient of acetonitrile. The gradient was formed with an aqueous 5% (v/v) solution of tetrahydrofuran as solvent A and acetonitrile as solvent B. Chromatography was performed at 55°C and the column was monitored at 269 nm.

2.2.14. Polyacrylamide gel electrophoresis

Denaturing polyacrylamide protein gels were run according to the method of Laemmli (1970) using the following stock buffer solutions: 30% acrylamide (acrylamide, 58 g; bisacrylamide, 2 g; made up to 200 ml with SDW and filtered through Whatman No.1 paper); separating gel buffer (Tris, 9.075 g; SDS, 2.5 g; made up to 50ml with SDW and adjusted to pH 8.8 with 10 N hydrochloric acid); stacking gel buffer (Tris, 3.025 g; SDS, 0.25 g; made up to 50 ml with SDW and adjusted to pH 6.8 with 10 N HCl); electrode buffer (Tris, 6.3 g; glycine 4.0 g; SDS, 1.0 g; made up to 1 l with SDW, pH should be approximately 8.3); ammonium persulphate (1% w/v in SDW); boiling mix (stacking gel buffer, 10 ml; SDS, 2g; β-mercaptoethanol, 5 µl; glycerol, 10 ml; bromophenol blue, 5 mg; made up to 100 ml with SDW).

12% gels were most commonly used to identify the pyruvate kinase

subunit (30% acrylamide stock, 2.78 ml; separating buffer, 1.75 ml; SDW, 2.06 ml; 1% ammonium persulphate, 0.38 ml). The gel was then polymerised by the addition of 20µl of TEMED (Sigma). 5.5% stacking gels were used (30% acrylamide, 0.55 ml; stacking gel buffer, 0.75 ml, SDW, 1.62 ml; 1% ammonium persulphate, 90 µl). The gel was polymerised with the addition of 10 µl of TEMED.

All polyacrylamide protein gels were run using the Hoefer 'Tall Mighty Small' minigel apparatus. Prior to use (to prevent protein contamination), the glass plates, aluminium back plates, formers and combs were washed with hot decon solution, rinsed with first hot water, then SDW and were then dried. Finally, all the equipment was wiped with a piece of cotton wool soaked in 100% ethanol. The gel sandwich was then assembled according to the manufacturers directions and fitted into the lower buffer tank. The separating gel was poured using a 1ml Gilson pipette and overlaid with SDW. After the gel had set, the water was poured off and the stacking gel layered on top with the pipette. A well-former was then added to the gel which was allowed to set for twenty minutes at room temperature, subsequent to which, the well-former was removed and the wells flushed out with a small volume of electrode buffer (to remove unpolymerised acrylamide). The minigel apparatus was assembled according to the manufacturers protocol and the buffer tanks filled with electrode buffer. Protein samples were prepared by addition of an equal volume of boiling mix and incubating in a boiling water bath for 5 min. After this denaturation step, the sample mix could be added to the well using a Hamilton microsyringe. After sample loading, the gel was run at 20 mA (200 V, 2 W power) for 1 hr or until the dye front had reached the bottom of the gel.

When the run was complete, the apparatus was disassembled and the gel sandwich was prised apart using a spacer. The gel was then ready for staining/destaining procedures.

2.2.15 Polyacrylamide gel staining and destaining

Coomassie blue staining/destaining: After running and removal from the gel apparatus, polyacrylamide gels were fixed in 20% (v/v) methanol, 10% (v/v) acetic acid for 15 min, stained in 0.25% (w/v) Coomassie blue R-

250, 45% (v/v) methanol, 9% (v/v) acetic acid for 20 min at room temperature, then destained in 5% (v/v) methanol, 10% (v/v) acetic acid for 1 hr at room temperature or until the background was sufficiently clear.

Silver staining/destaining : If greater sensitivity of staining was required, gels were first stained with Coomassie blue using the procedure above with the exception that the destaining buffer was 50% (v/v) methanol, 10% (v/v) acetic acid. Destaining was for 2 hr at room temperature and was followed by rinsing with 100 ml SDW for 10 min. The gels were then soaked in 100 ml silver stain solution (stain component A: silver nitrate, 0.4 g in 2 ml SDW; stain component B: 0.36% (w/v) sodium hydroxide, 21 ml, 14.8 M ammonium hydroxide, 1.4 ml; stain formed by adding A to B dropwise, stirring constantly and making the final solution up to 100 ml with SDW) for 15 min. The stain was removed and the gels washed in SDW for 5 min. After removal of the water, the gels were soaked in developer (1% (w/v) citric acid, 2.5 ml; 38% (v/v) formaldehyde, 0.25ml; made up to 500 ml with SDW) for 5-15 min or until band formation had occurred. Development was then stopped by transferring the gels to 200 ml of stop solution (50% (v/v) methanol; 10% (v/v) acetic acid).

If over-developing occurred (i. e. the background became too dark) the process could be reversed by using silver destain buffer which was made up as follows: 100ml destain X (NaCl, 3.7 g; $CuSO_4$, 3.7 g; SDW, 85 ml; add ammonia solution dropwise until the buffer turns deep blue and then make up to 100 ml with SDW) was added to an equal volume of destain Y (sodium thiosulphate, 43.6 g in 100 ml SDW) and the solution shaken until it became colourless. Over-stained gels were then added to this solution for 1-2 min or until the bands had faded sufficiently. The gels were then transferred to stop solution to arrest the destaining process.

Silver staining is thought to be between 10 and 100 fold more sensitive than Coomassie staining but is especially sensitive if the gel has been already stained with Coomassie blue R-250. This is thought to be due to the Coomassie blue providing extra nucleation sites for silver deposition within the protein bands.

2.2.16 Densitometry of polyacrylamide gels

Polyacrylamide gels stained with Coomassie blue were scanned with a Joyce Loebl Chromoscan 3 instrument fitted with a 100 W tungsten halogen lamp and an aperture of 5 x 0.5 mm. The results were processed with an IBM compatible computer running the 'scan' program (Chromoscan 3 manual revision 5.1) which integrated the data and performed peak spotting allowing quantitation of the protein bands.

2.2.17. Succinylation and clostripain digestion

Succinylation of the lysine residues of the purified pyruvate kinase was based on the method of Mitchell (1977).

Succinylation: 0.1 ml of 1 mg/ml pyruvate kinase was placed in a microcentrifuge tube together with 0.1 ml of 2 M Tris. The contents of the tube were mixed and a small amount of succinic anhydride was added every 30 min. The reaction was complete after 4 hr. The succinylated protein was then desalted by centrifugation through 1 ml of a Bio-Gel P-6DG column in a microcentrifuge tube. To prepare the columns 1 g of dry gel beads was added to 10 ml of water in a small flask, and the gel was allowed to hydrate for 4 hr at room temperature with occasional stirring. Most of the water was then decanted to leave a 50% slurry (enough for 12 columns). The column was prepared by punching a small hole in the bottom of a microcentrifuge tube with a hypodermic needle and placing a wad of glass wool over this hole. 1 ml of the Bio-Gel P-6DG slurry was then poured into the tube and the 'column' placed in a supporting plastic tube and centrifuged for 15 sec at 1000 r.p.m. in a bench centrifuge. 200 µl of the succinylated protein was pipetted onto the dried gel and the microcentrifuge tube transferred to another supporting tube and centrifuged again. The desalted protein pooled in the bottom of the supporting tube and was ready for clostripain digestion.

Clostripain digestion: Clostripain was first activated by preincubating the enzyme at a concentration of 2 mg/ml in calcium acetate (1 mM; DTT, 2 mM) overnight at 4°C. The succinylated protein, prepared as above (100 µl of the desalted sample; 0.05 - 0.5 mg protein), was placed into a microcentrifuge tube together with 10 µl of DTT (75 mM in 1% (w/v)

ammonium bicarbonate). After mixing, activated clostripain was added to the tube (5 µl or 0.01 mg clostripain), and the tube placed into a 37 °C water bath. Complete digestion occurred within 2-4 hr.

2.2.18. Cyanogen bromide cleavage

CNBr cleavage was performed according to the method of Gross and Witkop (1962): 25 µl of 1 mg/ml purified pyruvate kinase was placed in a microcentrifuge tube, to which was added 175 µl of formic acid, 70% (v/v). A small crystal of CNBr was then added and the tube covered in foil and left at room temperature for 24 hr. The cleaved protein was dissolved in 200 µl of SDW and was freeze dried.

2.2.19. Preparation of *Thermoplasma* genomic DNA

Genomic DNA from *T. acidophilum* cells was prepared by the method of Sambrook *et al.* (1990). 0.2 g (wet weight) of cells was resuspended in 3.2 ml of Tris/sucrose buffer (Tris HCl, 50 mM; sucrose, 25% (w/v); pH 8). Cells were lysed by the addition of 2 ml of SDS/EDTA buffer (SDS, 5% (w/v); EDTA, 0.15 M; pH 8) and the mixture swirled on ice for 5 min. After addition of 2.5 mg of proteinase K, the mixture was incubated at 45°C for 30 min. 1/10 volume of sodium acetate (3 M; pH 5.5) neutralised the mixture, which was then extracted with an equal volume of phenol/chloroform (1/1 v/v). After allowing the layers to separate, the upper layer was aspirated with a Pasteur pipette and two volumes of ice cold ethanol were added to precipitate the total nucleic acids. The resultant precipitate was spooled with a glass rod and resuspended in 5 ml of 2xSSC (NaCl, 0.3 M; sodium citrate, 0.03 M) followed by RNA digestion with the addition of 0.25 ml of RNase A (10 mg/ml) and incubation at 37°C for 20 min. A further phenol/chloroform extraction removed the digested RNA and the remaining nucleic acids (the DNA) were precipitated with 2.5 volumes of ice cold ethanol. The precipitate was pelleted by centrifugation in a bench microfuge at 10000g for 5 min, the pellet dried under vacuum and then resuspended in SDW.

2.2.20. Restriction digests of genomic DNA

Restriction digests of *T.acidophilum* genomic DNA were performed according to the method described in Hanish and McClelland (1988). 2 µl of 10x KGB buffer (potassium glutamate, 1 M; Tris acetate, 250 mM, pH 7.5; magnesium acetate, 100 mM; BSA fraction 5, 0.5 mg/ml, β-mercaptoethanol, 5 mM) were added to 18 µl of a solution of the genomic DNA as prepared in section 2.2.19. (containing approximately 25 ng of DNA). The solutions were briefly mixed and then 1 U of the appropriate restriction enzyme(s) was added. The mixture was incubated for 2 hr at room temperature and then stopped with the addition of EDTA (0.5 M, pH 7.5) to a final concentration of 10 mM. The digest was then ready to be analysed on an agarose gel as described in section 2.2.21.

2.2.21. Agarose gel electrophoresis.

Horizontal agarose DNA mini-gels were run according to the method described in Sambrook *et al.* (1990). The gel was composed of 1% (w/v) agarose (separating range 0.5-7.0 kb) in TBE buffer (Tris, 90 mM; boric acid, 90 mM; EDTA, 2 mM; ethidium bromide added to 0.5 mg/ml). The gel was poured using the Pharmacia horizontal mini-gel apparatus as per manufacturers instructions and the buffer tank set up containing approximately 350 ml of TBE buffer. 20 µl samples were loaded in solution with 6 µl of FBX loading buffer [Ficoll, 15% (w/v); bromophenol blue, 0.25% (w/v); xylene cyanol FF, 0.25% (w/v)]. If the gel was to be calibrated, one lane was reserved for 5 µl of Pharmacia λHinD III digest which contains DNA marker bands of the following sizes: 23.130 kb, 9.414 kb, 6.557 kb, 4.361 kb, 2.322 kb, 2.027 kb, 0.560 kb and 0.125 kb. DNA bands on the gel were visualised using a long wavelength UV transilluminator and, if necessary, the gels were photographed using Polaroid type 665 black and white film.

2.2.22. Southern blotting.

Transfer of DNA from agarose gels to nitrocellulose membranes was performed according to the method of Southern (1975). DNA was run out on an agarose gel as described in section 2.2.21. and then denatured by shaking in 200 ml of a solution of NaOH, 0.5 M and NaCl, 1.5 M for 1 hr. The gel was then neutralised by shaking in 200 ml of a solution of Tris-HCl, 1 M

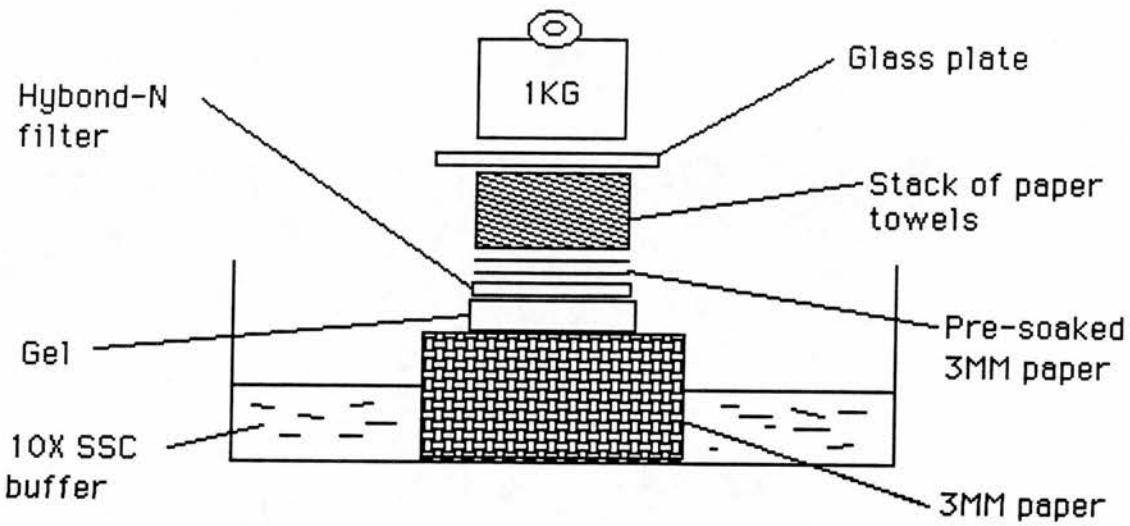


Figure 2.1: Apparatus used for Southern blotting.

and NaCl, 1.5 M (pH 8.0) for 2 hr with three changes of buffer. The Southern blot was then set up as in Figure 2.1.

The gel was surrounded by clingfilm so as to prevent 'short-circuiting' by the paper towels making contact with the soaked 3MM paper. The towels, the 3MM paper and the nitrocellulose filter (Hybond-N, Amersham) were cut to exactly the size of the gel and, for later orientation, the top left hand corner of the filter was nicked with a pair of scissors. 1 l of 10x SSC buffer was used and contained NaCl, 87.5 g and sodium citrate, 45.0 g.

Transfer was considered to be complete after leaving overnight. The apparatus was then disassembled and the DNA fixed to the membrane by exposure to UV light on a transilluminator for 3 min.

2.2.23. Non-radioactive labelling and detection of oligonucleotide probes.

Two separate systems for the non-radioactive labelling and detection of oligonucleotide probes were employed. The digoxygenin system of Boehringer-Mannheim involves the labelling of the oligonucleotide probes with the hapten digoxygenin (DIG). After hybridisation of the probes, they are detected with an anti-DIG antibody coupled to alkaline phosphatase which catalyses the production of a fluorescent compound which binds to the membrane. The ECL (enhanced chemiluminescence) system of Amersham International involves tailing the probes with fluorescein-labelled dNTP's and detection with an anti-fluorescein antibody coupled to horseradish peroxidase which, when incubated with the manufacturers detection solutions produces fluorescence which can be detected on X-ray film.

2.2.23.1. Labelling

The digoxygenin system (Boehringer Mannheim): The method involves the 3' end tailing, with terminal transferase, of the oligonucleotide with a homopolymer of dUTP, some of which are labelled with the hapten digoxygenin. The method described below is as set out in the manufacturers kit instruction manual. 140 pmol of the oligonucleotide probe (2 µl) was added to 4 µl of 5x tailing buffer (Tris-HCl, 15 mg/ml; cacodylate, 214 mg/ml; BSA, 1.25 mg/ml; pH 6.6), 2 µl $CoCl_2$, 1 µl DIG-dUTP

labelling mixture and 9 µl of SDW. The tailing reaction was then started with the addition of 2 µl of terminal transferase (55 U). The mixture was incubated at 37°C for 15 min and then placed on ice, the reaction being stopped by adding 1 µl of stop solution (glycogen, 0.1 mg/ml; EDTA, 74 mg/ml; pH 7.5). This gave a mixture of labelled oligonucleotide probe and non-incorporated nucleotides. These could be separated by the following method: the mixture of probe and non-incorporated nucleotides was added to 2.5 µl of lithium chloride solution (4 M) and 75 µl of chilled ethanol. The solutions were mixed and precipitated at -70°C for 30 min. The precipitate of labelled probe was collected by centrifugation at 13500 r.p.m. (10000g) in a bench microfuge for 15 min. It was then washed in cold ethanol, dried under vacuum and resuspended in SDW.

The enhanced chemiluminescence system (Amersham): This method also involves 3' tailing of the oligonucleotide probe with the enzyme terminal transferase. In this case, however, the label is not the hapten digoxygenin but fluorescein-11-dUTP. The method of labelling is as described in the manufacturers manual. 140 pmol of the oligonucleotide probe (2 µl) was added to 10 µl of fluorescein-11-dUTP, 16 µl of cacodylate buffer, 116 µl of SDW, 16 µl of terminal transferase. The solutions were mixed and then incubated at 37°C for 90 min. In this case, there is no need to purify the labelled probe from the non-incorporated nucleotides as they will be washed away during the hybridisation/detection procedure.

2.2.23.2.. Detection: use of the oligonucleotide probes.

The digoxygenin system (Boehringer Mannheim): The oligonucleotide probe in use was labelled and purified as described in section 2.2.23. and the Southern blot prepared as in section 2.2.22. The hybridisation procedure was adapted from Schmitz *et al.*, (1990): The nitrocellulose filter was first prehybridised in a hybridisation bag for 3 hr at 52°C in 20 ml of hybridisation buffer (0.5% (w/v) blocking reagent, 0.1% (w/v) N-laurylsarcosine, 0.02% SDS, made up to 100 ml with 5x SSC, pH 7.0). This buffer was discarded and replaced with 2.5 ml/100 cm^2 of membrane of hybridisation buffer containing 140 pmol of labelled oligonucleotide probe. Hybridisation was for 16 hr with constant gentle distribution of the

solution.

The hybridisation temperature was set at 5°C below the melting temperature (Tm) of the oligonucleotide in use. Individual Tm values were calculated using the formula (Bolton and McCarthy, 1962):

$$Tm = 81.5 - 16.6(\log_{10}[Na^+]) + 0.41(\%G+C) - (600/N)$$

where N is the length of the oligonucleotide.

After hybridisation, the membranes were washed twice at room temperature with 50 ml/100 cm^2 of 2X SSC, 0.1% (w/v) SDS. This was followed with a stringent wash with 0.1X SSC, 0.1% (w/v) SDS at a temperature between 52 and 68°C depending on the length and sequence of the probe. The stringent wash was repeated once and subsequently, the membranes used for detection of hybridised nucleotides. The following procedure describes detection of a membrane of 100 cm^2 area using the fluorescent substrate AMPPD for alkaline phosphatase: All incubations were at room temperature and with gentle shaking. A 1 min wash of the membrane in 100 ml of buffer 1 (Tris-HCl, 100 mM; NaCl, 150 mM; pH 7.5) was followed by an incubation for 30 min in the same volume of blocking buffer (blocking reagent, 1% (w/v) in buffer 1). The blocking buffer was then replaced by 20 ml of a solution of freshly diluted <DIG>:AP (1:5000; 150 mU/ml in blocking buffer) and incubation proceeded for a further 30 min. The membrane was then washed twice with 100 ml of washing buffer (Tween 20, 0.3% (v/v); maleic acid, 11.61 g/l; NaCl, 8.77 g/l) for 15 min to remove unbound antibody. Eventually, the membrane was equilibrated for 2 min with 20 ml of buffer 2 (Tris-HCl, 100 mM; NaCl, 150 mM; $MgCl_2$, 50 mM; pH 9.5). The AMPPD stock solution was diluted (100 µl in 10 ml) and the membrane was incubated in 10 ml of the diluted solution for 5 min. The membrane was then blotted on 3MM paper (not to complete dryness), sealed in a hybridisation bag, pre-incubated for 15 min at 37°C and then exposed to X-ray film (Hyperfilm-ECL, Amersham)in a Dupont Cronex cassette. The exposure time depended on the strength of the fluorescent signal and ranged from 2 to 24 hr.

The ECL system (Amersham): The basic hybridisation protocol used

was very similar to that used with the digoxygenin system. The membrane was prehybridised at 42°C for 30 min in hybridisation buffer (Hybridisation buffer component, 0.1% (w/v); Blocking reagent, 0.5% (w/v); SDS, 0.02% (w/v) in 5X SSC). Labelled probe was then added to the buffer to a concentration of 5-10 ng/ml and the membrane hybridised at 5°C below the Tm of the probe for 2 hr in a shaking water bath. The membrane was washed for 5 min in 5X SSC, 0.1% (w/v) SDS twice followed by a more stringent wash in 1X SSC, 0.1% (w/v) SDS at 42-50°C for 15 min. The membrane was then incubated at room temperature for 1 min in buffer 1 (NaCl, 150 mM; Tris base, 100 mM; pH 7.5) and blocked with 0.5% (w/v) blocking agent in buffer 1 (0.25 ml of buffer/ cm^2 of membrane). The blot was again rinsed with buffer 1 for 1 min and then incubated with 25 ml of the antibody conjugate solution (anti-fluorescein-HRP conjugate diluted 1:1000 in buffer 2 - NaCl, 400 mM; Tris base, 100 mM; pH 7.5 - containing 0.5% (w/v) BSA fraction V) for 30 min. The membrane was then rinsed with buffer 2 for 5 min three times to remove unbound antibody. Equal volumes of the two detection solutions were mixed (0.125 ml/cm^2 of membrane) and the membrane incubated at room temperature in the mixture for 1 min. Excess mixture was drained off, the blot wrapped in cling-film and exposed to Hyperfilm-ECL in a Dupont Cronex cassette for 15 min to 3 hr depending on the strength of the signal.

Hyperfilm-ECL results from both detection systems were processed automatically using a Fuji RG-11 X-ray film processor.

2.2.24. Polymerase Chain Reaction

Amplification of part of the pyruvate kinase gene, and part of the Ta20Kp gene, was performed using the polymerase chain reaction (PCR) as described in Sambrook *et al.* (1990). Three reactions were used, the first with probe 1 (sense) mixed with probe 2 (antisense), the second with probe 2 (antisense) and probe 3 (sense) and the third with probe 1 (sense) and probe 4 (antisense). Probes 1 and 2 were specific for the *T. acidophilum* pyruvate kinase sequences, probe 3 for the Ta20KP sequence and probe 4 was specific for another conserved pyruvate kinase sequence (but not one obtained from the *T. acidpohilum* enzyme). The reaction mixture was as follows: 10 µl, 10X PCR buffer (Tris-HCl, 100 mM; $MgCl_2$, 15

mM; KCl, 500 mM; gelatine, 1 mg/ml; pH 8.3); 10 µl, 2 mM dNTP's; 2 µl, oligonucleotides (1 mg/ml); 5 µl, template DNA (25 ng); 0.4 µl, *Taq* polymerase (1000-5000 U/ml) made up to 100 µl with SDW. The mixtures (in 0.5 ml eppendorf tubes) were then placed in a Techne PHC-2 thermal cycler and subjected to 30 cycles of the type: denaturation (95°C, 1 min), annealing (45-55°C, 1 min) and extension (72°C, 3 min). Annealing at lower temperature allowed non-specific priming for the TA20KP reaction whilst the higher temperature improved the specificity of priming for the pyruvate kinase reactions. The generated PCR products were then either used as hybridisation probes or were cloned for sequencing experiments.

2.2.25. DNA sequencing

Sequencing reactions were performed using either the 'fmol DNA sequencing system' from Promega or the Taq DyeDeoxy Terminator cycle sequencing kit from Applied Biosystems.

(a) fmol cycle sequencing kit (Promega)

The following method of direct incorporation is as described in the technical manual accompanying the kit. For each set of sequencing reactions, four 0.5 ml microcentrifuge tubes were labelled G, A, T or C. Each of the four d/ddNTP mixtures was prepared by adding 1 volume of the supplied nucleotides to 1 volume of SDW. 1 µl of the appropriate diluted mix was added to each tube and the tubes stored at 4°C until needed. For each set of four reactions, the following reagents were mixed in a microcentrifuge tube: 1 µg of template DNA; 25 ng of sequencing primer; 0.5 µl of radiolabel ([α-^{35}S]dATP or [α-^{32}P]dATP or [α-^{33}P]dATP); 4.25 µl of 5X fmol sequencing buffer (Tris-HCl, 250 mM; $MgCl_2$, 10 mM) and SDW to a final volume of 16 µl. 1 µl of sequencing grade *Taq* DNA polymerase (5 U/ml) was added to the mixture and mixed in by pipetting up and down. 4 µl of this enzyme/primer/template mix was then added to the inside wall of each tube containing the d/ddNTP mix followed by 20 µl of mineral oil. The tubes were then spun briefly in a microcentrifuge. The reactions took place in a Techne PHC-2 thermal cycler which was preheated to 95°C for 2

min followed by 30 cycles of denaturation (95°C, 30 sec); annealing (42°C, 30 sec) and extension (70°C, 1 min). After the thermal cycling finished, 3 µl of fmol sequencing stop solution was added to each tube and the tubes spun briefly in a microcentrifuge to terminate the reactions. The contents of the tubes were then ready for loading onto a sequencing gel.

(b) Taq dyedeoxy sequencing (Applied Biosystems)

The principle behind this approach to thermal cycle sequencing is the same as that for the fmol kit described above. The DyeDeoxy kit is, however, optimised for a different detection system - the Applied Biosystems Model 373A DNA sequencing instrument. The performance of the kit relies on four dye-labelled dideoxynucleotides: G, A, T and C DyeDeoxyterminators. These are used in place of standard dideoxynucleotides in the enzymatic sequencing reaction and incorporate a fluorescence-based dye into the DNA along with the terminating base. All four termination reactions are performed in a single eppendorf tube and hence the chances of tube-handling/pipetting errors occurring are much reduced. The kit utilises dITP (which minimises band compressions) and the thermally stable AmpliTaq DNA polymerase. For each reaction, the tube contained: template DNA, 2.5 µg; primer, 50 ng (together in a total volume of 10.5 µl); 5XTACS buffer (400 mM TrisHCl, 10 mM $MgCl_2$, 100 mM $(NH_4)_2SO_4$, pH 9), 4 µl; dNTP mix (750 µM dITP, 150 µM dATP, 150 µM dTTP, 150 µCTP), 4 µl; 1 µl each of G, A, T and C DyeDeoxy terminators and 0.5 µl of the AmpliTaq enzyme. The contents of the tube were overlaid with 20 µl of paraffin oil prior to thermal cycling (initial heat denaturing to 96°C for 2 min followed by 25 cycles of 96°C for 30 s, 50°C for 15 s and 60°C for 4 min). After completion of the cycling reactions, the reaction mix was made up to a volume of 100 µl with SDW and removed from underneath the oil. The termination products of the sequencing reactions were then separated from non-incorporated nucleotides by two extractions using an equal volume of phenol:water: chloroform (68:18:14) at room temperature. The purified extension products were then precipitated by adding 0.3 volumes of 3 M sodium acetate, pH 5.5 and 2.5 volumes of absolute ethanol. After incubating at -80°C for 1 hr the mixture was centrifuged for 15 min at room temperature and the resulting pellet washed with 70% ethanol and

dried in a Speedvac. The pellet was then dissolved in 4 µl of deionised formamide and, prior to loading on a gel, heated to 90°C to denature the double-stranded products.

Sequencing gels

Sequencing gels were prepared according to the method described in Sambrook *et al.* (1990). Glass plates were thoroughly cleaned before assembling the gel sandwich so as to prevent the formation of air bubbles on pouring the gel. The gel was composed of 75 ml of 6% acrylamide/urea top solution [75 ml, 40% acrylamide solution (acrylamide 380 g; N,N'-methylene bis acrylamide, 20 g; made up to 600 ml with SDW); 50 ml, 5X TBE; 230 g, urea; adjusted to 500 ml with SDW]; 0.42 ml of 10% (w/v) ammonium persulphate solution; and was polymerised with the addition of 40 µl of TEMED. The gel was then quickly poured in one motion, the shark's tooth comb inserted invertedly and the gel allowed to set for 45 min. After setting, the gel was pre-run at 55 W (1600 V) power for 45 min in electrophoresis buffer (0.5X TBE).

For the fmol sequencing kit, the sequencing reaction tubes were heated to 70°C for 2 min and 3 µl of each was loaded, in the order TCGA, into the spaces formed by the comb, now inserted the correct way up. The gel was then run at 55 W (1600 V) power until the marker dye had migrated to about three quarters of the way down the gel. The gel sandwich was prised apart using a spatula and the gel fixed in 50% (v/v) methanol, 10% (v/v) acetic acid for 15 min on its supporting glass plate. After fixing, the gel was transferred to a piece of 3MM paper which had been precisely cut to the dimensions of the gel. The gel was dried in a vacuum drier for 1 hr and then exposed to X-ray film (Hyperfilm) in a Dupont Cronex cassette for 16-24 hr depending on the strength of the signal. The film was then developed automatically in a Fuji RG-11 X-ray film processor.

For the Taq DyeDeoxy kit, the gels were prepared analogously and the samples were loaded after the gel had been inserted into the Applied Biosystems 373A DNA sequencer. After the gel had been running at 30 W for 10 min, the laser scan was initiated and data collection commenced. Sequence raw data obtained was processed and enhanced using software packages supplied with the instrument.

2.2.26 Extraction of DNA from agarose

DNA fragments were extracted from agarose gels using a simple electroelution device constructed according to the method of Pai and Bird (1991). Bands of interest - PCR products for example - were run out on horizontal agarose gels and then excised from the gel as small slices of agarose. A strip of plastic (2 x 7 x 0.4 cm) was cut to fit a submarine gel electrophoresis tank sideways. Two holes were punched into the strip 1 cm apart and 0.6 cm from the top. A blue (1 ml) disposable pipette tip (Sarsten) was loaded with 0.4 ml of molten 0.6% agarose in 1XTAE (0.04 M tris-acetate and 0.002 M EDTA, pH 7.8) which was allowed to set, forming a plug in the tapered end of the tip. The agarose gel slice(s) containing the bands of interest were then loaded onto the solidified plug and surrounded in molten agarose to a depth of 1-2 mm above the top of the slice(s). When the agarose had set, the tip was cut at a distance of 2 cm from the tapered end with a sharp scalpel whilst the top was filled with 1XTAE and covered in a small piece of dialysis membrane. Another blue pipette tip was cut at 3.2 cm from the tapered end and pushed into the top of the first tip, thus sealing the dialysis membrane. The cut end of the agarose containing tip was then fitted into one of the holes in the plastic screen and the entire apparatus submerged in the buffer tank containing 1XTAE running buffer with the screen facing the cathode. A current of 0.15 A was then applied to the tank and the progress of the DNA fragments through the gel could be monitored if ethidium bromide was introduced into the buffer. Typically, after 30 min, all of the DNA had been released into the compartment of buffer between the agarose plug and the dialysis membrane. This buffer could then be collected and, if desired, concentrated for further work.

2.2.27 Ligation of DNA

With both vectors, ligations were carried out using a 1:1 molar ratio of vector to insert. For example, 27.8 ng of the 0.6 kb PCR product generated using probes O1 and O4 was ligated with 50 ng of the pCR II TA cloning vector (Invitrogen) in a final volume of 11 µl. Ligation buffer contained 50 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 20 mM DTT, 1 mM ATP, 100 mg/ml BSA and 1 unit of T4 ligase. The reaction was incubated at $12^{o}C$ overnight.

2.2.28 Preparation and transformation of competent *E. coli* cells

A single colony of *E. coli* TG1 cells from a stock minimal media plate was used to inoculate 10 ml of 2X YT which was incubated overnight at 37 ^{o}C. 0.5 ml of overnight culture was then used to inoculate 50 ml of 2X YT broth which was grown at 37 ^{o}C in an orbital incubator to an OD at 600 nm of 0.4-0.5. The YT broth was then chilled on ice for 10-15 min. Cells were pelleted by centrifugation (3000 g at 4^{o}C for 5 min) and resuspended in 16.5 ml sterile 50 mM $CaCl_2$ solution. After incubation on ice for 30 min, the cells were again pelleted by centrifugation (2000 g at 4 ^{o}C for 5 min) and resuspended in 4 ml sterile 50 mM $CaCl_2$. Optimum transformation frequencies were obtained using freshly-made competent cells although they could be stored at -70 ^{o}C in a 3:1 culture:glycerol ratio solution.

50 µl of competent cells were thawed on ice for each ligation reaction. 2 µl of 0.5 M β-mercaptoethanol was added to each tube of cells followed by 1µl of each ligation reaction and the contents of the tubes were then mixed by gentle tapping. The tubes were incubated on ice for 30 min. and then for 60 seconds in a 42^{o}C water bath. After this incubation, the tubes were replaced on ice for 2 min. and then 1 ml of LB (for pBluescript II ligations) or SOC (for pCRII reactions) was added to each tube. The cells were then incubated at 37^{o}C for 1 hour at 225 r.p.m. in a gyratory shaker-incubator. Whilst the tubes were shaking, LB agar plates containing 50 µg/ml Kanamycin were prepared by spreading 25 µl of X-Gal (40 mg/ml stock solution) and 25 µl of IPTG (40 mg/ml stock solution) on top of the agar with an L-shaped glass spreader.

After 1 hr, the tubes containing the transformed cells were placed on ice and then 25 µl of each was spread onto a different LB agar plate using the spreader. The plates were inverted and left overnight in a 37^{o}C oven. The following day the plates were examined for growth and discrete white colonies were selected for plasmid isolation.

2.2.29 Colony lifts

Colony lifts were carried out as described below, according to the method of Grunstein and Hogness (1975): Hybond N membrane was cut to the size and shape of the plates to be blotted, marked with a non-symmetrical nick pattern and then lowered onto the agar surface. After 1 min., the membrane was removed and placed colony-side-up on a piece of filter paper which had been previously soaked in denaturing solution (1.5 M NaCl, 0.5 M NaOH). After 7 min., the membrane was transferred to a piece of filter paper which had been soaked in neutralising solution [1.5 M NaCl, 0.5 M Tris-HCl (pH 7.2), 1 mM EDTA]. After 3 min., the membrane was transferred to fresh pads, again soaked in neutralising solution, for a further 4 min. The membrane was washed in 2X SSC and then dried at room temperature. The DNA was fixed to the membrane either by baking at 80°C for 2 hr or by exposure to UV on a transilluminator for 5 min. The colony blot was then prehybridised, hybridised and washed as described for Southern blots (see 2.2.23).

2.2.30 Rapid preparation of plasmid DNA using Promega 'Magic Minipreps'

Cells from an overnight culture (1.5 ml) of the TG1 strain containing the desired plasmid were harvested by centrifugation (5000 g for 10 min. at 4°C) and resuspended in 200 ml of 'cell resuspension solution' [50 mM Tris-HCl (pH 7.5), 10 mM EDTA, 100 mg/ml RNase A]. The resuspended cells were transferred to a microcentrifuge tube to which was added 200 µl of 'cell lysis solution' (0.2 M NaOH, 1% SDS). The contents of the tube were mixed by inverting until the cell suspension had cleared. Subsequently, 200 µl of 'neutralisation solution' (2.55 M KOAc, pH 4.8) were added to the tube, the contents of which were again mixed by inversion. The tube was then centrifuged at 14000 g for 5 min. The supernatant from this spin was decanted into a fresh microcentrifuge tube to which was added 1 ml of 'Magic DNA purification resin'. The contents of the tube were mixed by inversion for 1 min. and then the resultant slurry pushed into a 'Magic mini-column' using a 5 ml syringe (the column was attached to the luer-lok extension of the syringe barrel).

The mini-column was then washed with 2 ml of 'column wash solution'

[200 mM NaCl, 20 mM Tris-HCl (pH 7.5), 5 mM EDTA - diluted 1:1 with absolute ethanol] using the 5 ml syringe. The washed column was transferred to a fresh microcentrifuge tube and centrifuged for 20 seconds to dry the resin. The mini-column was transferred to another microcentrifuge tube and then the plasmid DNA was eluted from it by applying 50 µl of pre-heated (70°C) TE buffer [10 mM Tris-HCl (pH 7.5), 1 mM EDTA] to the column and centrifuging for 20 seconds. The mini-column could then be discarded and the eluent in the microcentrifuge tube containing the plasmid could be stored at -20°C prior to further use.

2.2.31 Matrix-assisted laser desorption/ionisation mass spectrometry

A protein of unknown function, purified from *T. acidophilum* using part of the protocol developed for purification of the pyruvate kinase (see appendix 1), was subjected to mass analysis in a LASERMAT Mass Analyser (Finnigan Mat). From SDS-PAGE analysis, the relative molecular mass of the protein was estimated to be approximately 20K.

20 µl of a diluted sample of the purified protein (containing 10 pmol of the protein) was loaded onto the matrix disc of the sample holder designed to fit the machine. The sample was loaded into the machine as per the manufacturer's instruction booklet and subjected to 11 laser shots at an accelerating voltage of 20018 V, the spectra of which were summated until a clear peak resulted. The machine was externally calibrated using a cytochrome c calibration file for determination of the mass of the protein.

Chapter 3: Growth and Identification of *T. acidophilum*

CHAPTER 3: GROWTH AND IDENTIFICATION OF *THERMOPLASMA ACIDOPHILUM*

3.1 Introduction

This chapter describes the methods used, and the results obtained, for the growth of *T. acidophilum* cells. It also describes the means used to confirm the identity of the cells. These included analyses of morphology by electron microscopy, and of enzyme content by assay of the archaebacterial isocitrate dehydrogenase.

3.1.1 Growth of *T. acidophilum* cells

The unusual (and extreme) growth conditions of the organism necessitated the construction of a special batch facility for their culture. This growth chamber used is illustrated in Fig. 3.1 and consisted of a 2l Pyrex glass flask with a fitted lid possessing four 'quick-fit' extensions. The extensions were occupied by (a) an aeration supply, (b) a refluxer, (c) a stirring paddle and (d) a tube containing a temperature probe. The aeration supply was a peristaltic pump which was set up to produce a steady air-flow of 100ml/min through the growth medium. Since the growth conditions of the organism are so harsh ($60^{o}C$, pH 2), the likelihood of contamination was considered to be low and hence no elaborate precautions (other than the use of filters) were taken to exclude atmospheric microorganisms from the culture apparatus.

The constant air-flow through the medium, coupled with the high temperature, meant that a significant reduction in the batch volume was a risk. This was averted by the use of the refluxer which allowed constant air flow but prevented loss of fluid from the medium by evaporation. The contents of the flask were mixed using a stainless steel paddle of the highest quality (this was required to prevent rusting in the extreme conditions within the flask) and the temperature was monitored using a temperature probe connected to a microprocessor controlled thermostat. This system maintained the temperature of the flask at $60^{o}C$ ($\pm 1^{o}C$). Continuous culture of the cells was propagated by replacing the cell-containing medium with fresh medium without washing out the flask in between. The small residual volume of cells remaining in the flask was enough to seed the second batch of medium for cell growth. Growth of

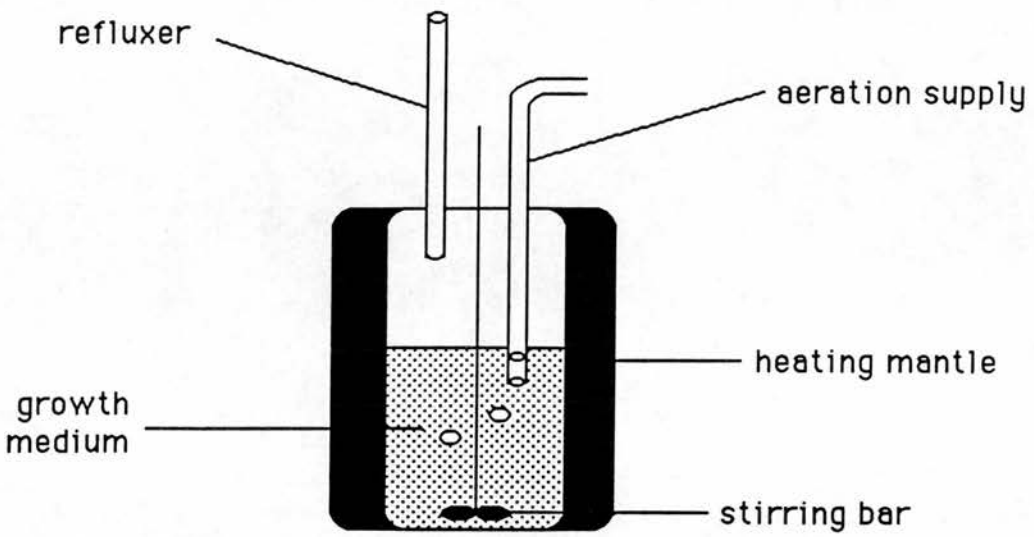


Figure 3.1: Apparatus used for continuous batch culturing of *T. acidophilum* cells. The growth medium was that described in section 2.2.1. All metal components in contact with the medium were of highest grade stainless steel.

the cells was monitored by frequent sampling of the medium and reading the absorbances of those samples at 650 nm.

3.1.2 Identification of the cultured cells

Although no contamination of the cells was anticipated, the cultured cells were frequently checked for homogeneity under the light microscope and their fine structure determined with an electron microscope. The size, morphology and general subcellular organisation of *T. acidophilum* had been previously determined (Darland et al., 1970) and hence any cells not conforming to the published parameters for the archaebacterial cells would be contaminants. The cells were also occasionally grown in the presence of the antibiotic vancomycin which was known not to affect *T. acidophilum* growth but which will prevent the growth of other thermophiles.

The archaebacterial nature of the cultured cells was assessed by examining the kinetic properties of the enzyme isocitrate dehydrogenase. This enzyme catalyses the oxidative decarboxylation of isocitrate to 2-oxoglutarate. Eukaryotes contain two activities of this enzyme. The first is an NAD^+-linked enzyme, confined to the mitochondria, which is allosterically regulated and has a controlling role in ATP generation by the citric acid cycle. The second is $NADP^+$-linked and is not allosterically regulated. In contrast, most bacteria have only the $NADP^+$-specific, non-allosteric enzyme (Weitzman, 1981). Danson and Wood (1984) reported finding both isocitrate dehydrogenase activities present in the sulphur-dependent, thermoacidophilic archaebacterium *Sulfolobus solfataricus*. Crucially, it appeared that both activities were functions of the same enzymic protein. As part of the identification of the cultured cells in this study, therefore, the isocitrate dehydrogenase activity of the cells was examined with respect to its cofactor specificity. The results of this analysis are discussed later.

3.2 Results

3.2.1 Growth of *T. acidophilum* cells

The growth curve for the cells is shown in Fig. 3.2. Previous work on this organism had shown that its temperature and pH optima for growth were 59°C and pH 2.0 respectively and hence, these were the

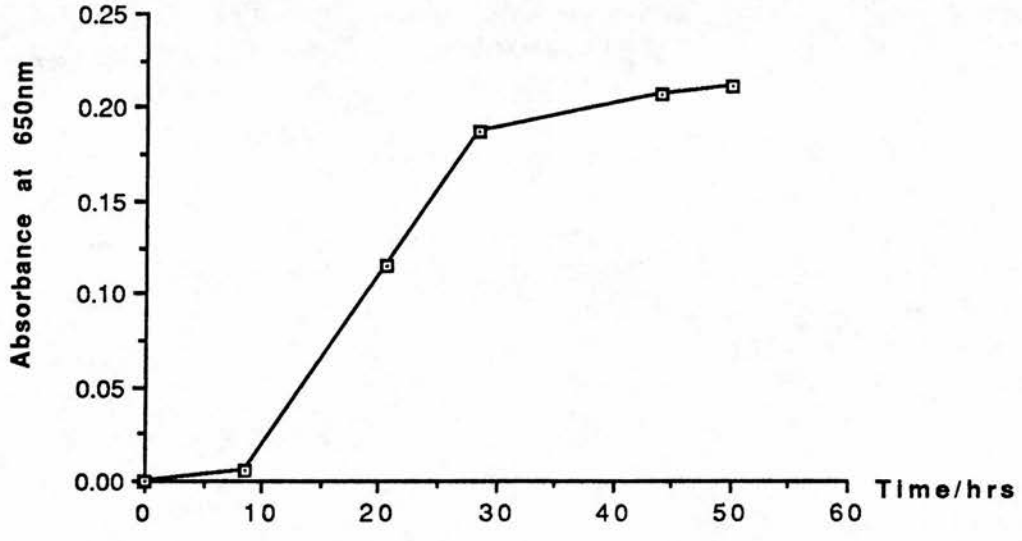


Figure 3.2: Growth curve for *T. acidophilum* cells. The culture was grown in the standard growth medium at pH 2 and 60^{o}C. The cell growth was followed by OD measurements (of 1 ml samples) at 650 nm.

parameters chosen for batch culture experiments. The growth curve shown in Fig. 3.2 has three distinct phases. There is a latent phase of about 10 hr duration which is followed by an exponential phase which continues up to about 45-50 hr. After this point, growth slows into the stationary phase during which cells are harvested. The maximum cell density reached corresponded to approximately 0.3 absorbance units at 650 nm. On harvesting, this gave 400 mg of cells from the 2l culture. The doubling time of the organism is about 4 hr during the exponential phase of growth. All of these values are typical of *T. acidophilum* cells (Darland *et al.*, 1970).

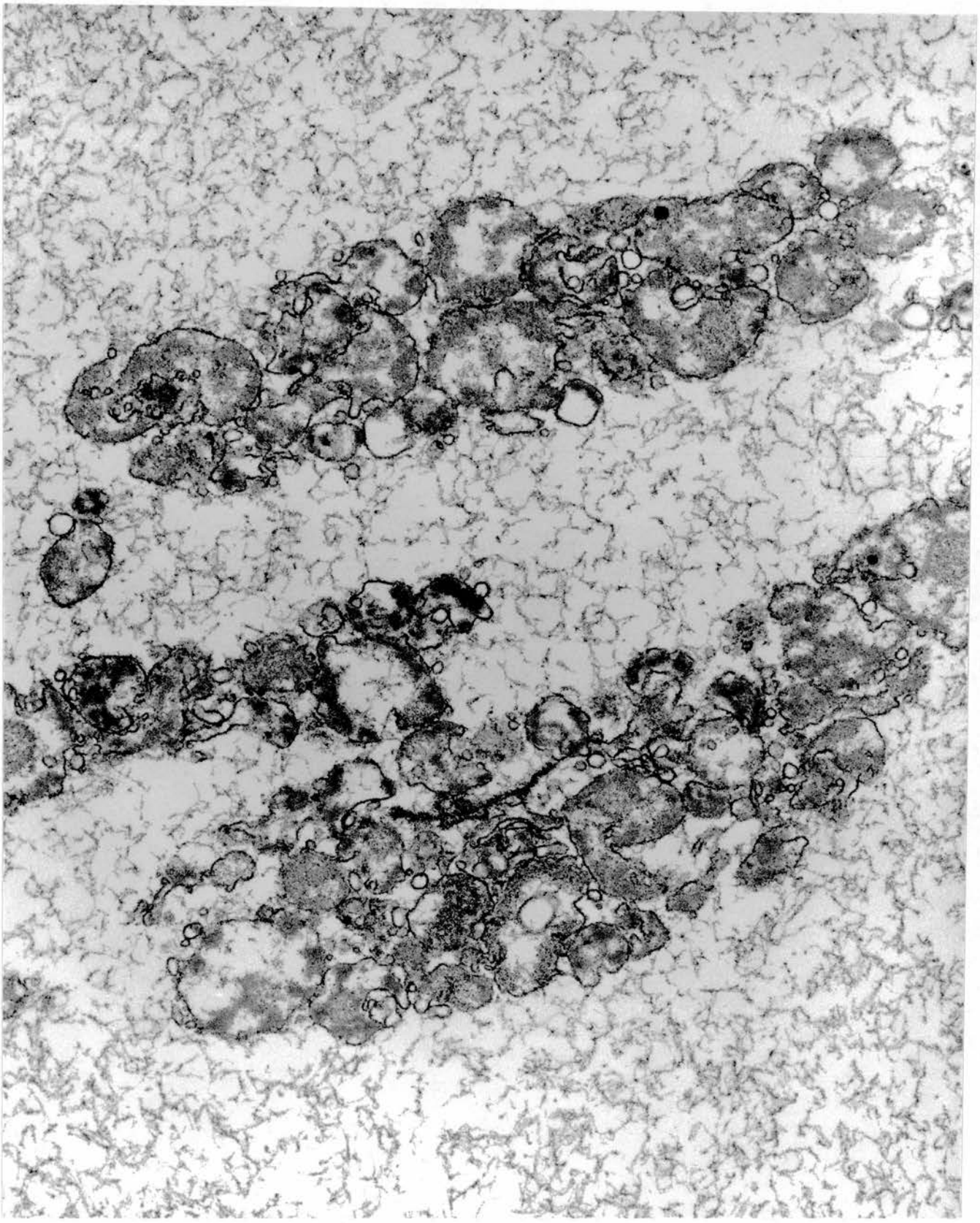
3.2.2 Electron microscopy of the cultured cells

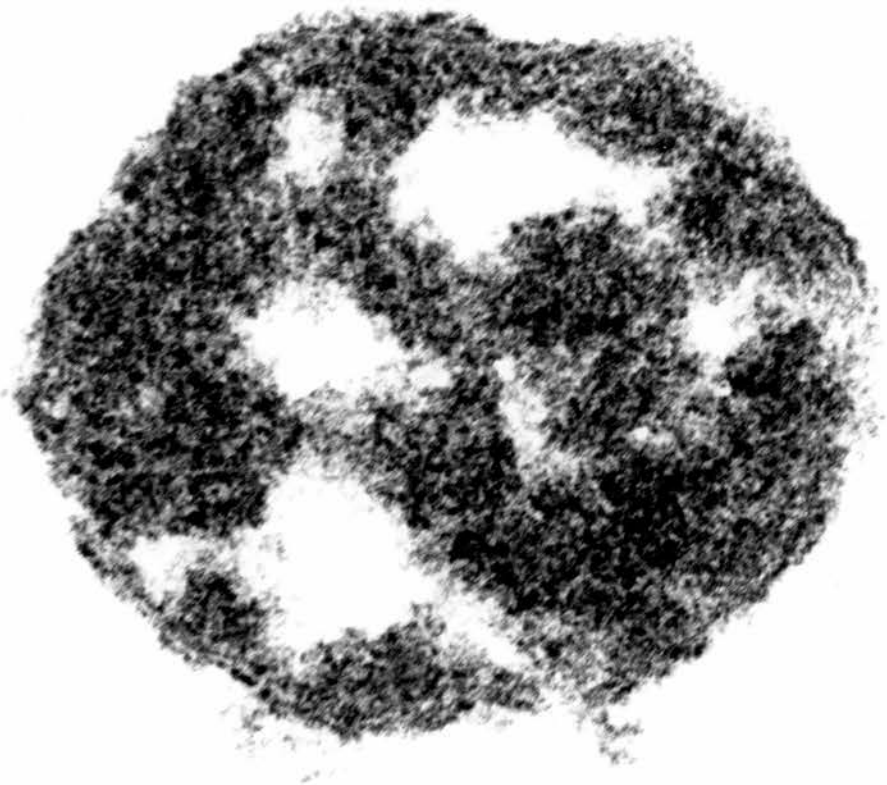
The cells were prepared for electron microscopy according to the method described in 2.2.2 and the resultant micrographs are shown in Figs. 3.3-3.5. Fig. 3.3 shows a grouping of cells in a clump. This is a characteristic arrangement of *T. acidophilum* cells and stands in contrast with Sulfolobus cells, which exist singly, and *Pyrodictium* cells, which form networks connected by fibres (Cowan, 1992). The plates in Figs. 3.4 and 3.5 show single cells which appear, morphologically, to be spherical in nature. They vary in size between 0.2 and 1.5 μm in diameter.

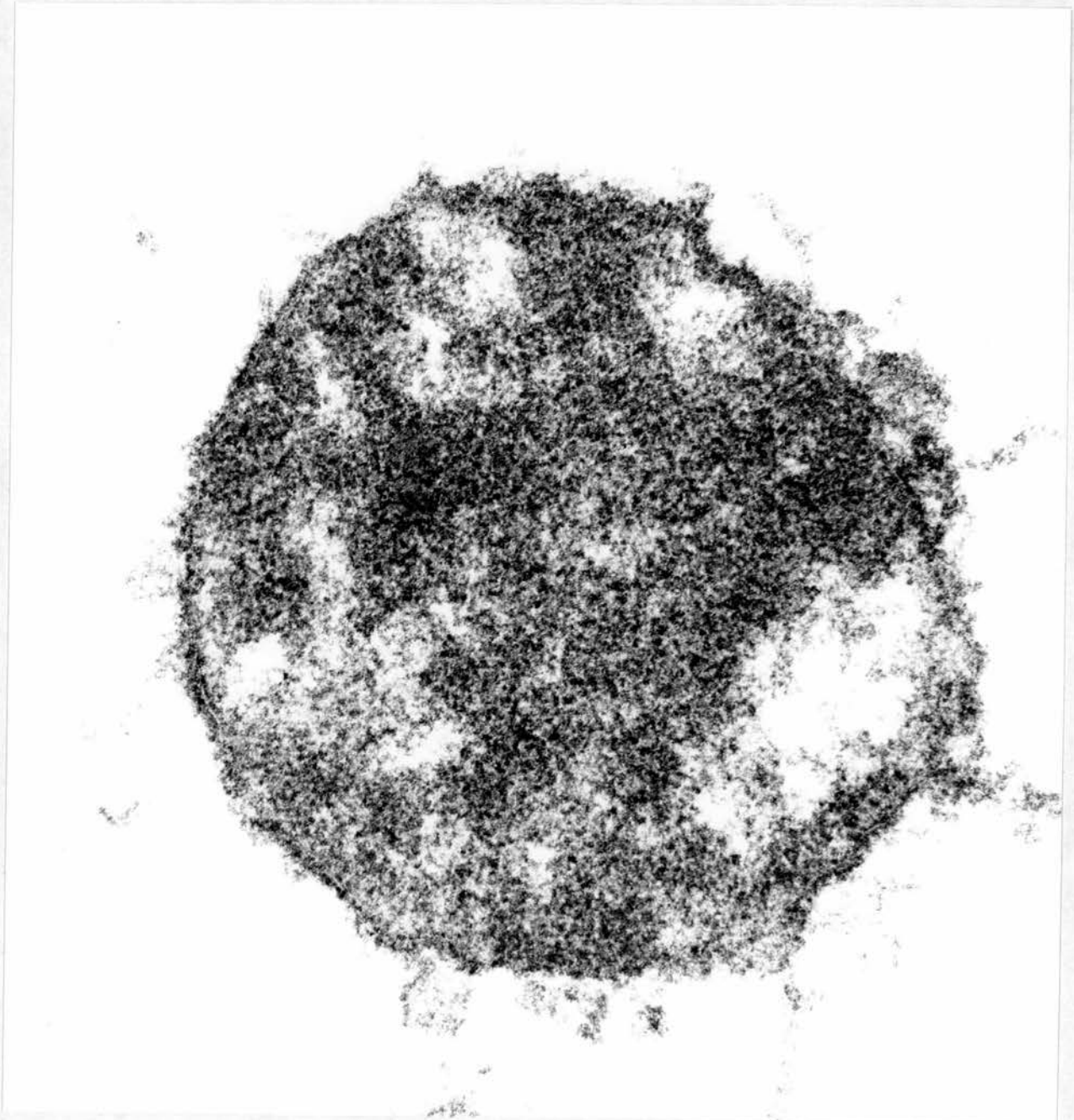
The micrographs reveal the relatively simple structure of a prokaryotic organism. There is no evidence of a limiting nuclear membrane and hence the nuclear material is likely to be dispersed throughout the cytoplasm. Equally, there is no evidence of any membranous organelles within the cytoplasm. Unlike bacterial cells, however, they lack a rigid cell wall and are separated from the surrounding environment only by a double membrane. This is evidenced by the tripartite appearance of the membrane structure. The absence of any cell wall structure was obliquely confirmed by two pieces of evidence. First, the addition of SDS to a preparation of the cells results in rapid lysis. Secondly, cell growth proved to be insensitive to the antibiotic vancomycin [an inhibitor of cell wall synthesis which blocks the addition of the muramic acid-lipid complex to an acceptor (Matsuhashi *et al.*, 1965)] even at concentrations of up to 5 mg/ml. The antibiotic is known to be active under the growth conditions used as it has been previously used to prevent the growth of other thermoacidophilic isolates (Darland *et al.*, 1970).

Figure 3.3: Transmission electron micrograph of a clump of *T. acidophilum* cells. Each cell measures between 0.5 - 3.0 μm in diameter. The cells were stained with uranyl acetate and osmium tetroxide.

Figures 3.4, 3.5: Transmission electron micrographs of single *T. acidophilum* cells. Worthy of note are the lack of any cell wall and the partially visible tripartite membrane structure (10 - 12 nm thick).







3.2.3 Isocitrate dehydrogenase assay of cell contents (Potter, 1993)

The crude extract of the cultured cells (prepared as in 2.2.3) was assayed for the presence of isocitrate dehydrogenase activity under two sets of conditions (2.2.8): (a) using 0.2 mM $NADP^+$ as cofactor, and (b) using 1.0 mM NAD^+ as cofactor. The reactions, in final volumes of 1 ml each, were started with the addition of the crude extract and was followed by the increase in A_{340}. Both $NADP^+$ and NAD^+-linked activities were detected in the cellular extract.

The activities depended hyperbolically on the concentrations of the cofactors (Figs. 3.6 and 3.7). For the NAD^+-dependent activity, the Km^{app} = 6.6 mM (+ 0.6 mM) and for the $NADP^+$-dependent activity, the Km^{app} = 32 µM (+ 4 µM). For each cofactor, the Vmax was similar [3.23 (+ 0.03) µmol/min for the NAD^+-linked activity and 3.14 (+ 0.04) µmol/min for the $NADP^+$-linked activity]. These values are broadly in agreement with those found for the enzyme isolated from *S. acidocaldarius* (Danson and Wood, 1984).

Assays of isocitrate dehydrogenase were also carried out using NAD^+ and $NADP^+$ together. The results of three assays, one with NAD^+ alone, one with NADP+ alone and one in the presence of both cofactors, are tabulated in Table 3.1. The activities noted for each cofactor were not additive, indicating that they are not the result of two separate enzymes. In fact, the results suggest that the two cofactors compete with one another, as was found to be the case for the *S. acidocaldarius* activities. They do so with Ki values approximately equal to their Km values - an observation which suggests that both cofactors bind at the same site on the enzyme. The theoretical enzyme activity for one enzyme capable of utilising both cofactors was calculated according to the equation (Segel, 1975):

$$V_{TOTAL} = \frac{V_{NAD}(1+[NAD]/Km^{NAD}) + V_{NADP}(1+[NADP]/Km^{NADP})}{1+[NAD]/Km^{NAD} + [NADP]/Km^{NADP}}$$

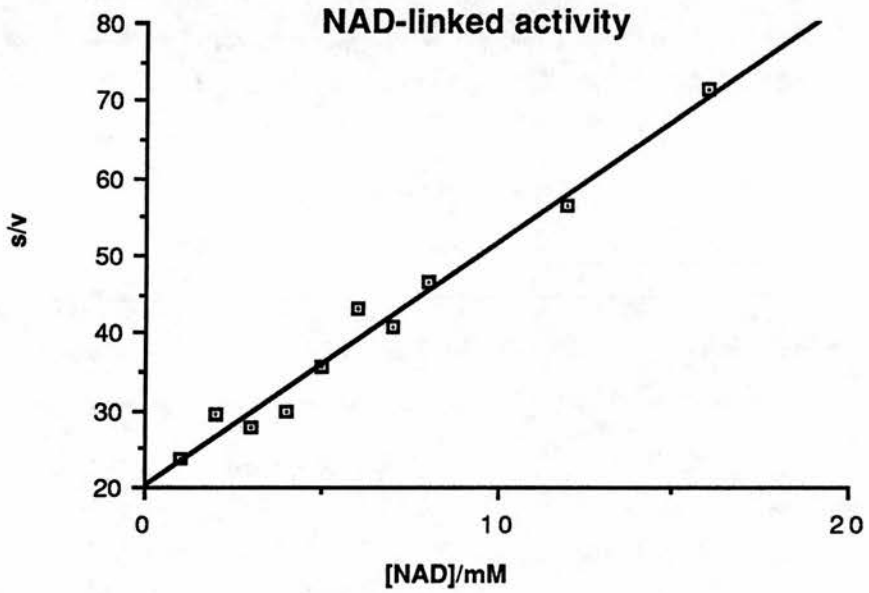


Figure 3.6: Hanes plot of dependence of the archaebacterial isocitrate dehydrogenase activity on [NAD^+]. The straight line fit of the data confirms that the enzyme is hyperbolically dependent on the concentration of that cofactor.

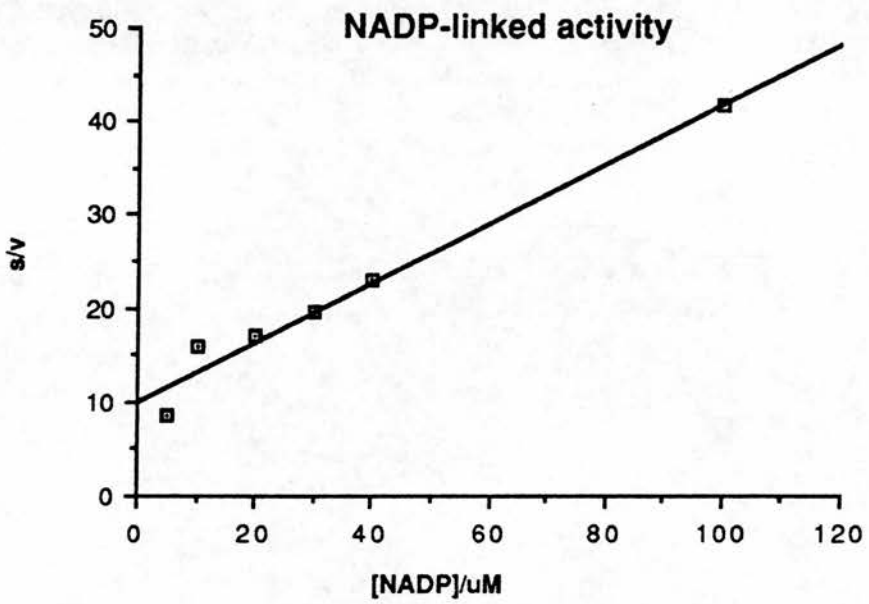


Figure 3.7: Hanes plot of dependence of the archaebacterial isocitrate dehydrogenase activity on [$NADP^+$]. The straight line fit of the data confirms that the enzyme is hyperbolically dependent on the concentration of that cofactor.

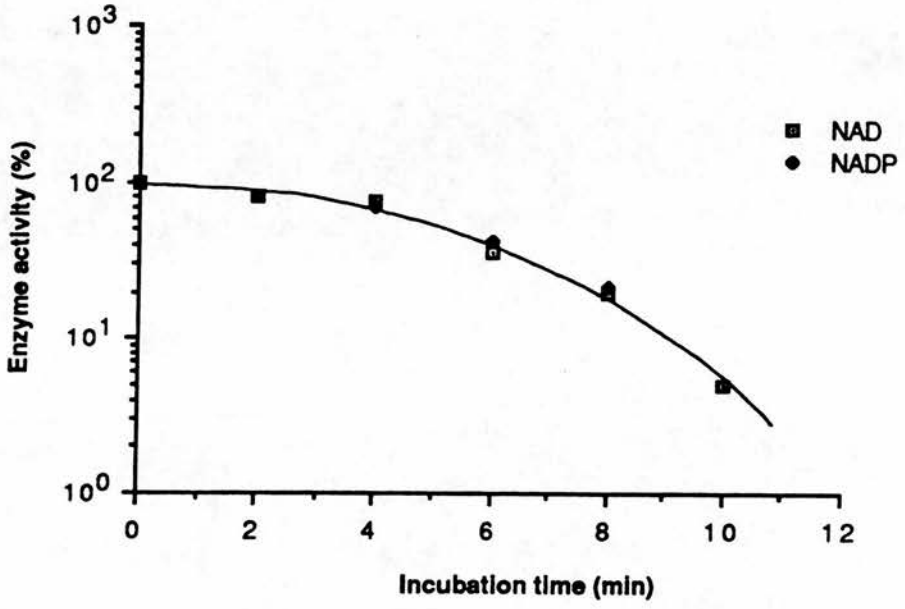


Figure 3.8: Thermal inactivation of the isocitrate dehydrogenase activities. Crude extracts of the cells were treated as described in the text and the resulting activity was tested with both cofactors at known time intervals.

substrate	activity		theoretical activity
	T. acido	*S. acido*	
1.0 mM NAD^+	0.17	0.19	
0.2 mM $NADP^+$	0.94	0.94	
1.0 mM NAD^+ + 0.2 mM $NADP^+$	0.90	0.91	0.94

Table 3.1: Substrate competition between cofactors for *T. acidophilum* isocitrate dehydrogenase. Theoretical activity for one enzyme capable of using both cofactors was calculated from Danson and Wood (1984). Enzyme activity is in units of μmol product/min/ml. *T. acido, Thermoplasma acidophilum; S. acido, Sulfolobus acidocaldarius.*

where: V_{NAD} and V_{NADP} are enzyme velocities in NAD^+ or $NADP^+$ alone and the Km values are as above.

Additional evidence for the presence of a dual-specificity isocitrate dehydrogenase enzyme was derived from a thermal inactivation study of the activities. Crude extracts of the cultured cells were incubated at 90°C and samples were removed at known time intervals, allowed to cool to the assay temperature and then assayed using the procedure described in 2.2.8. The results of the experiment are shown in Fig. 3.8 - both activities are lost coincidentally. It was surprising that the rate of inactivation increased with time and may be due to problems in maintaining the high temperatures involved. The implications of these results are discussed in the following section.

3.3 Discussion

The cells cultured in this study grew well at pH 2.0 and at 60°C with a doubling time of 4 hr and to a maximum density of 0.28 absorbance units at 650 nm. Electron microscopy of the cells indicated that they were prokaryotic in organisation, did not possess cell walls and tended to associate together in 'clumps'. The cell morphology was variant but generally spherical in nature and they varied in size from 0.2 - 1.5 mm. All of these parameters are typical of *Thermoplasma acidophilum* cells as documented in the literature. The kinetic evidence presented indicates that, like *Sulfolobus acidocaldarius*, the cultured cells possess an isocitrate dehydrogenase which can utilise both NAD+ and NADP+ as cofactors (since the km for NAD+ is 200 fold higher than that for NADP+, it is possible that the use of the former is not physiologically significant, although many enzymes have been observed with weaker NADH/NADPH specificities). All of this evidence taken together suggests that the cultured cells are indeed *T. acidophilum* and hence all subsequent work is relevant to that organism.

The kinetic evidence presented for the isocitrate dehydrogenase is doubly interesting. First, if the conclusion drawn from the data is a correct one, then an isocitrate dehydrogenase exhibiting dual specificity for nicotinamide cofactors may be a widespread phenomenon throughout the archaebacteria, as *T. acidophilum* and *S. acidocaldarius* are from different branches of the archaebacterial phylogenetic tree. This would be in contrast with the bacteria, which in most cases have only the

$NADP^+$-linked enzyme and with the eukaryotes, which have both activities but as functions of separate enzyme proteins. This observation may help to confirm the status of the archaebacteria as a separate taxonomic 'domain' of organisms (the 'archaea' - Woese *et al.*, 1990) and to refute other work which splits up the archaebacteria, combining some with the eukaryotes and others with bacteria (Lake, 1989).

Secondly, several other groups have observed archaebacterial proteins which display dual specificity towards NAD^+ and $NADP^+$ cofactors: glyceraldehyde-3-phosphate dehydrogenase (Fabry and Hensel, 1987), glucose dehydrogenase (Danson, 1988) and malate dehydrogenase (Grossebuter *et al.*, 1986). Whilst the phenomenon is certainly not specific to the archaebacteria, it does appear to be more prevalent amongst their enzymes than amongst the comparable enzymes from both eukaryotes and bacterial sources. It is, therefore, possible that the detection of such dual-specificity enzymes may assist in the classification of newly discovered organisms (Fuhrman *et al.*, 1992).

A summary table of the results discussed in this chapter can be seen in Table 3.2.

feature	found in cultured cells	characteristic of *Thermoplasma*
opt. temp.	59°C	✔
opt. pH	2.0	✔
doubling time	4 hr	✔
max. A_{650}	0.26 AU	✔
prokaryotic	✔	✔
cell groups	clumps	✔
cell wall	none, tripartite membrane	✔
morphology	variant	✔
cell size	0.3 - 2.0 microns	✔
IDH activity	both NAD and NADP cofactors used by the enzyme	✔(presumed)

Table 3.2: Characteristics of the cultured cells.

Chapter 4: Purification and Properties of Pyruvate Kinase from *T. acidophilum*

CHAPTER 4: PURIFICATION AND PROPERTIES OF PYRUVATE KINASE FROM *T. ACIDOPHILUM*

4.1 Introduction

This chapter describes the isolation of pure pyruvate kinase from *T. acidophilum* by a novel series of chromatographic steps, and its subsequent characterisation. Although many purification schemes for pyruvate kinases from various sources have previously been reported, they were deemed to be unsuitable for the archaebacterial enzyme for a number of reasons (which are discussed later). Hence, a new procedure which was faster and produced better yields from low amounts of starting material was designed. A means of stabilising the enzyme activity was also required as it proved to be susceptible to proteolysis and cold denaturation.

Having purified the pyruvate kinase to homogeneity, it was next important to gain some information about its physical and kinetic properties for comparison with mesophilic pyruvate kinases. The subunit molecular weight of the protein was determined under denaturing conditions. The dependence of the activity of the enzyme on the concentrations of its various substrates was investigated as were its regulatory properties. To complete the kinetic characterisation of the enzyme, the dependence of activity on pH and temperature were studied. A summary table of the properties of the *T. acidophilum* enzyme can be found at the end of the chapter. For comparative purposes, the table also includes the physical and kinetic characteristics of pyruvate kinases isolated from both a eukaryotic and a bacterial source.

4.2 Results 1: Purification

Many purification protocols have been published for pyruvate kinases isolated from various eukaryotic and bacterial sources. Amongst the most recent are those for the pyruvate kinases of the eukaryote Saccharomyces cerevisiae (Murcott *et al.*, 1991) and for the thermophilic bacterium *Bacillus stearothermophilus* (Sakai *et al.*, 1986). The former method was based on the affinity of the yeast enzyme for the matrix Cibacron blue 3G-A, a broad specificity matrix which binds many proteins including over 50 nucleotide requiring enzymes, albumin and interferon. This procedure was deemed unsuitable for purification of the *T.*

acidophilum enzyme for two reasons. First, the matrix was time-consuming to prepare and, secondly, the yield from the column was poor (McNally, 1991). The method for the yeast enzyme has since been adapted (Collins et al., 1992), abandoning the Cibacron blue step for a multi-column series which more closely resembles the protocol developed by Sakai and co-workers for the *Bacillus* enzyme. It was this adapted method which was used as the basis for the procedure developed for the archaebacterial enzyme. The finalised protocol is fully described in section 2.2.4 and begins with a cell lysis followed by an ammonium sulphate precipitation. These initial steps are the prelude to a series of chromatographic separations which resolve the archaebacterium's complement of proteins by size-exclusion, ion-exchange and affinity techniques. The protocol is summarised in Table 4.1.

The purification of *T. acidophilum* pyruvate kinase presents several specific problems. First, the maximum cell density of the archaebacterium is low (OD at 650 nm of 0.3 which corresponds to 200 mg cells/l) and hence the starting amounts of protein are necessarily low (it was discovered early on that the cells cannot be stored frozen without loss of their pyruvate kinase activity - this problem has since been countered with 'Yun buffer', the use of which is described later). Secondly, the pyruvate kinase, in common with its mesophilic counterparts from other organisms, appears to be particularly unstable. Other pyruvate kinases have been found to be susceptible to both proteolysis at exposed surface residues and to cold denaturation (Kuczenski and Suelter, 1972). The latter problem is thought to be due to the increase in concentration of the enzyme during the freezing process (Franks, 1991). An alternative explanation for the phenomenon depends on the buffer used to store the enzyme at sub-zero temperatures. If an enzyme is dissolved in a sodium phosphate buffer prior to freezing, the precipitation of NaH_2PO_4 (if this were the less soluble salt) which would occur during the freezing process would cause a rise in the pH of the remaining solution. Similar effects will be obtained with many other buffer systems. The pyruvate kinase of *T. acidophilum* is typical in that it too is susceptible to both proteolytic and cold denaturation processes, particularly the latter.

The problem of proteolysis of the enzyme was overcome in the initial stage of the purification protocol by including a cocktail of protease inhibitors (see 2.2.3) in the mixture used for lysis of the archaebacterial cells. To prevent cold denaturation of the enzyme, this and all the

STEP	ACTIVITY (U)	PROTEIN (mg)	SPECIFIC ACTIVITY (U/mg)	PURIFICATION
CELL-FREE EXTRACT	11.2	28.0	0.40	1
AMMONIUM SULPHATE PRECIPITATION	13.5	17.2	0.78	1.96
MONO-Q	14.0	0.56	25.0	62.5
SUPEROSE 6	13.8	0.28	49.5	124
5'AMP-AGAROSE	10.5	0.05	201.2	503

Table 4.1: Purification table for *T. acidophilum* pyruvate kinase

Units of specific activity are μmol NADH converted min^{-1} (mg of protein)$^{-1}$

Final recovery (from maximal activity found) is 75%.

subsequent steps were performed at 45-50°C, a temperature at which the enzyme was found to be both active and stable. In the case of the chromatographic purification steps, this necessitated the construction of column jackets (which are not commercially available for the columns used) composed of glass tubes wound into tight coils which were then slotted around the columns. Heated water was then pumped through the coils. Further, the buffers passing through the columns were maintained at 50°C in water baths and the temperature of each column eluent was monitored continuously until a steady temperature of 45-50°C was reached; only then was the sample applied to the column. After elution, fractions collected in microcentrifuge tubes were maintained at 50°C prior to assay for pyruvate kinase activity.

The choice of column matrices was based largely on the adapted protocol of Collins *et al.* (1992) although some changes were made. The Sephacryl S-200 size exclusion and Q-Sepharose anion-exchange steps were replaced by Superose 6 and Mono-Q columns simply because the latter columns are available on the Pharmacia FPLC system. The use of this system greatly accelerated the purification, reducing overnight column steps to twenty minutes each, with no loss of efficacy.

During initial purifications, a hydrophobic interaction step was used to complete the purification of the pyruvate kinase as the eluent from the Superose 6 column was not homogeneous. Pooled fractions containing pyruvate kinase activity from the Superose 6 column were loaded onto an Aquapore RP-300 column (7 μm particle size; 2.1 mm x 30 mm) attached to an Applied Biosystems 130A microbore separation system. The column was eluted with a linear 8-80% (v/v) gradient of acetonitrile in aqueous 0.1% TFA (concentration of TFA was constant throughout). Elution of the column was followed at 220 nm. Whilst this step yielded a homogeneous sample of protein, as evidenced by SDS-polyacrylamide gel electrophoresis (Fig. 4.1), the harsh elution conditions abolished the activity of the protein and hence it could not be used for further characterisation studies. Protein prepared in this way was, however, used for the amino acid analysis studies presented in chapter 5.

For studies requiring active protein, an affinity step proved to be a useful replacement for the final hydrophobic interaction step, yielding a

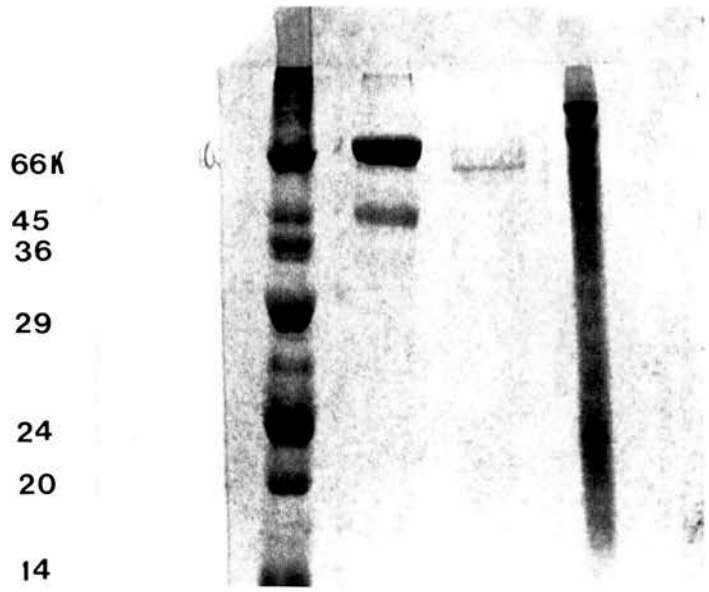


Figure 4.1: SDS-polyacrylamide gel of *T. acidophilum* pyruvate kinase purified using the destructive final H.P.L.C. step. The lanes contain (L-R): Sigma molecular weight markers (the topmost band is 60 k), Sigma rabbit muscle pyruvate kinase (the contaminating lower molecular weight band is thought to be enolase), *T. acidophilum* pyruvate kinase, *T. acidophilum* crude extract. Silver staining confirmed the homogeneity of the preparation.

homogeneous preparation of active protein. Fig 4.2a shows an SDS-polyacrylamide gel of the purified enzyme whilst Fig. 4.2b shows a densitometer scan of the gel. Cibacron blue was not used for the reasons outlined earlier and was, therefore, replaced by 5'AMP agarose (Sigma), a far more specific affinity matrix (it shows a narrower specificity for NAD^+-dependent dehydrogenases and ATP-dependent kinases). The purified protein obtained from this step was then stabilised using 'Yun buffer' (Yun *et al.*, 1976 and section 2.2.5). The buffer is essentially a phosphate buffer which contains glycerol (which prevents denaturation of the protein at surfaces) and β-mercaptoethanol (which reduces disulphides). Pyruvate kinase could be stored in this medium at room temperature or below for over 250 hr without loss of activity. Fig. 4.3 shows a comparison of two different preparations of the pyruvate kinase stored at 4^oC. One was stored in 'Yun buffer' and the other was stored in assay buffer. Whilst the former sample proved to retain all of its activity for the duration of the test, the sample stored in assay buffer had lost over half of its activity after 100 hr and no activity was detectable after 240 hr.

4.3 Results 2: Characterisation

4.3.1 Determination of relative molecular mass of pyruvate kinase subunits

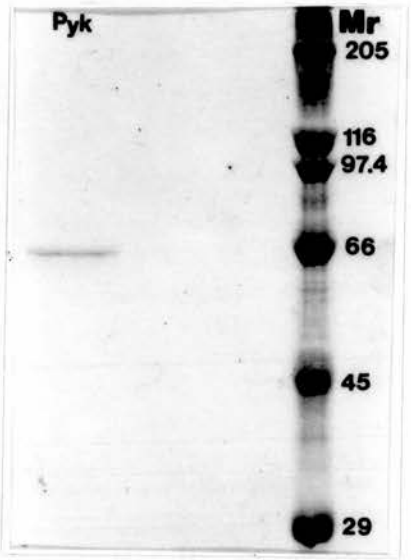
Fig. 4.2a shows a denaturing SDS-polyacrylamide gel of purified pyruvate kinase eluted from the final affinity column. By comparison with the running positions of marker proteins, the relative molecular mass (Mr) of the subunits of the protein was estimated to be 60k (Figure 4.4).

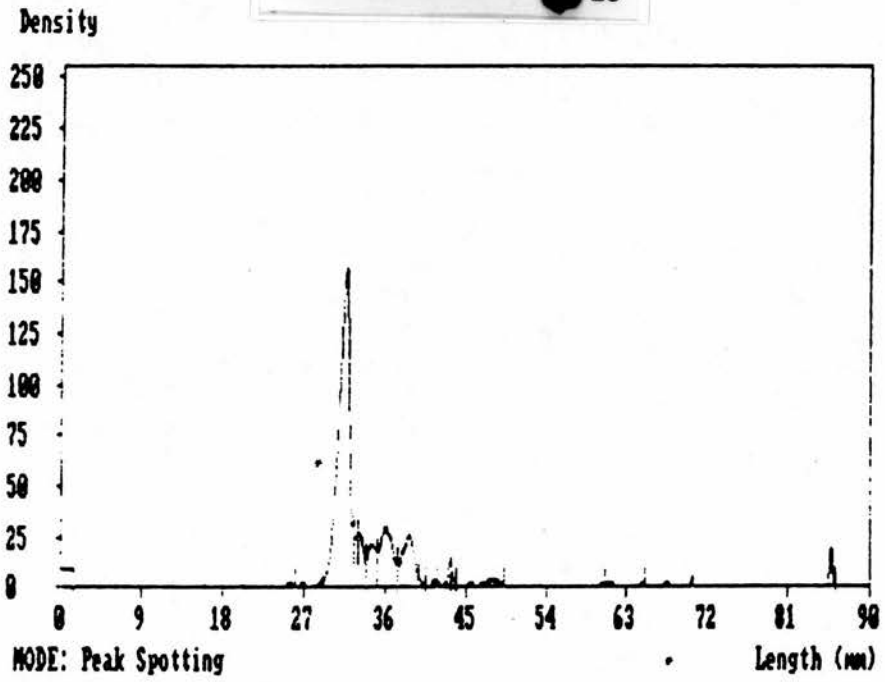
4.3.2 Kinetic characterisation

For kinetic studies, the coupled assay of Bucher and Pfleiderer (1955 and section 2.2.6.1) could not be used as the properties of the coupling enzyme, lactate dehydrogenase, would also be affected by changing conditions of the assays. This might, therefore, have interfered with the interpretation of assay results despite the fact that LDH is used in large excess. Hence, the direct spectrophotometric assay of Pon and Bondar (1967 and section 2.2.6.2) was used, at $60\,^oC$, for the kinetic

Figure 4.2(a): SDS-polyacrylamide gel of *T. acidophilum* pyruvate kinase purified using the non-destructive final affinity column step. The lanes contain (L-R): *T. acidophilum* pyruvate kinase, Sigma high molecular weight markers. Silver staining confirmed homogeneity of the preparation.

Figure 4.2(b): Densitometer scan of the polyacrylamide gel shown in Fig. 4.2(a). The scan was performed using a Joyce Loebl Chromoscan 3 instrument using an aperture of 0.5x5 mm. The scan estimated the pyruvate kinase to be 96% pure after background correction.


Pyk
Mr
205
116
97.4
66
45
29


Density
250
225
200
175
150
125
100
75
50
25
0
0
9
18
27
36
45
54
63
72
81
90
MODE: Peak Spotting
Length (mm)

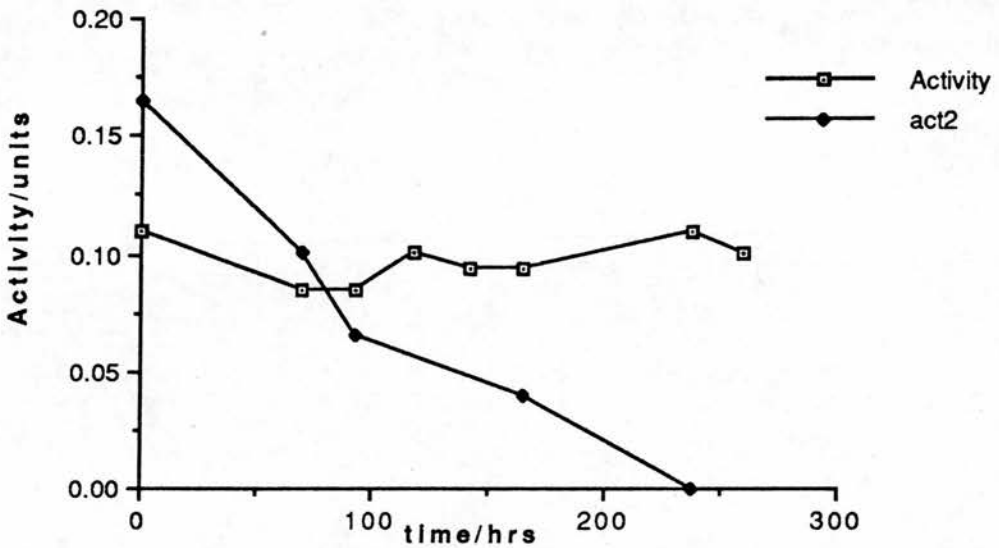


Figure 4.3: Comparative stability of *T. acidophilum* pyruvate kinase activity resuspended in assay buffer (activity) and in 'Yun buffer' (act2). Both buffers were maintained at $4^{o}C$. The activity of the enzyme was recorded as change in absorbance units at 340 nm/min.

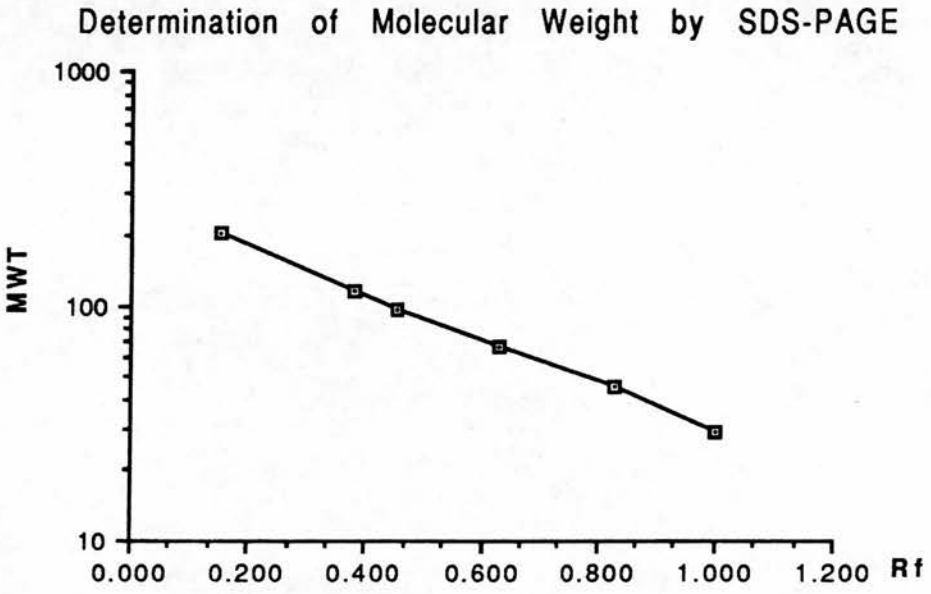


Figure 4.4: Plot of log Mr against distance moved through the SDS-polyacrylamide gel for the purified pyruvate kinase subunits. Points shown are distances moved by the standard proteins.

characterisation of the enzyme. This was considered to be a safer option for obtaining reliable results.

4.3.2.1 Dependence of activity on pH

The pH profile of the enzyme is shown in Fig. 4.5. The bell-shaped curve obtained is typical of mesophilic pyruvate kinases, as is the estimated pH optimum of 7.5. This result is not particularly surprising since *T. acidophilum* is known to maintain approximately neutral pH intracellularly despite the extremely acidic nature of its growth environment.

4.3.2.2 Dependence of activity on temperature

The dependence of the activity of the pyruvate kinase on ambient temperature was also investigated. A temperature profile for the activity is shown in Fig. 4.6. Initial rates are expected to increase exponentially with temperature (considering the principles of thermodynamics) until denaturation becomes so rapid that the initial rate cannot be measured. Ordinarily, curves of the type shown in Fig. 4.6 indicate that the increase in activity of the enzyme being studied is in linear relation to the increase in temperature. This situation carries on up to an 'optimum' temperature beyond which the relationship does not hold and the enzyme is starting to denature. The relationship of activity to temperature for an enzyme at temperatures below the 'optimum' is usually defined as the Q_{10} of the enzyme. The Q_{10} value for an enzyme, which is the ratio of the enzyme's activity at two temperatures which are 10 degrees apart, is usually between 1 and 2 (i.e. a 10 degree increase in ambient temperature typically results in no more than a 2 fold increase in enzyme activity). This type of relationship appears to hold for the *T. acidophilum* pyruvate kinase.

An Arrhenius plot for the pyruvate kinase is shown in Fig. 4.7 and a plot of A_{280} variation with temperature is shown in Fig. 4.8. The Arrhenius plot, of log k_{cat} (estimated from Vmax values for different initial concentrations of enzyme) for an enzyme against 1000/ temperature (in K), is typically a straight line from which the activation energy of the enzyme can be determined (slope of line = -Ea/2.303R, where Ea = activation energy and R = gas constant). For the *T. acidophilum* pyruvate kinase, a possibly biphasic plot is obtained from

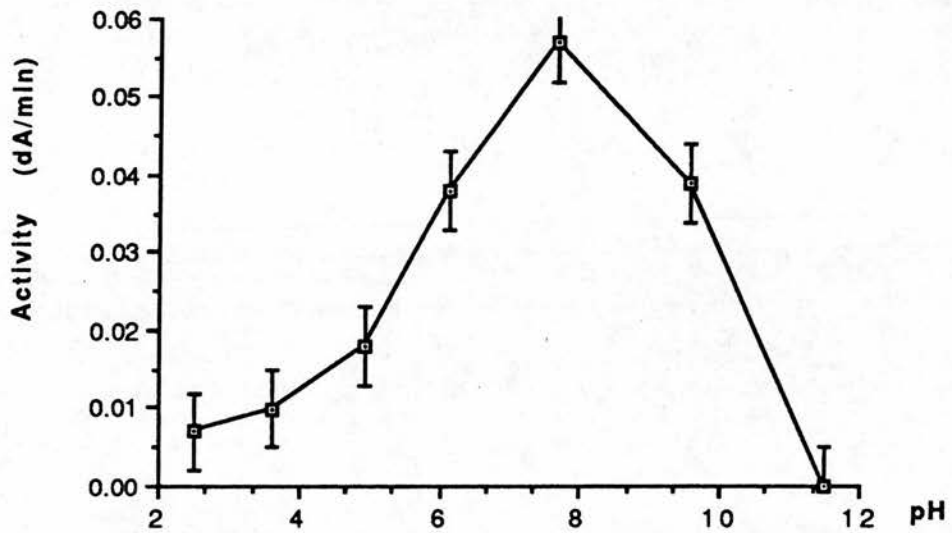


Figure 4.5: Dependence of *T. acidophilum* pyruvate kinase activity on pH. The activity of the enzyme was determined at 60°C using the direct spectrophotometric assay of Pon and Bondar (1967). Each determination was performed in triplicate.

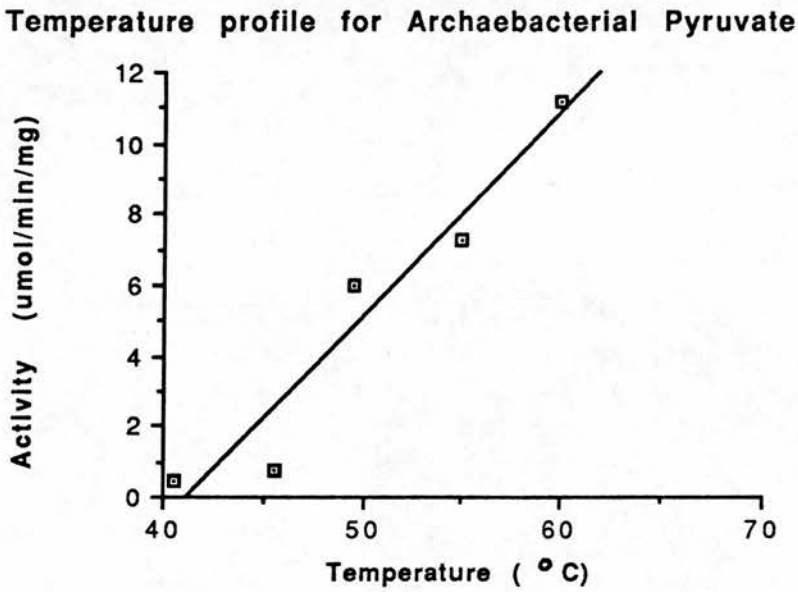


Figure 4.6: Dependence of *T. acidophilum* pyruvate kinase activity on temperature. The activity of the enzyme was determined using the direct spectrophotometric assay of Pon and Bondar (1967). The pH of the assay buffer was adjusted at each temperature to 7.5. Each determination was performed in triplicate.

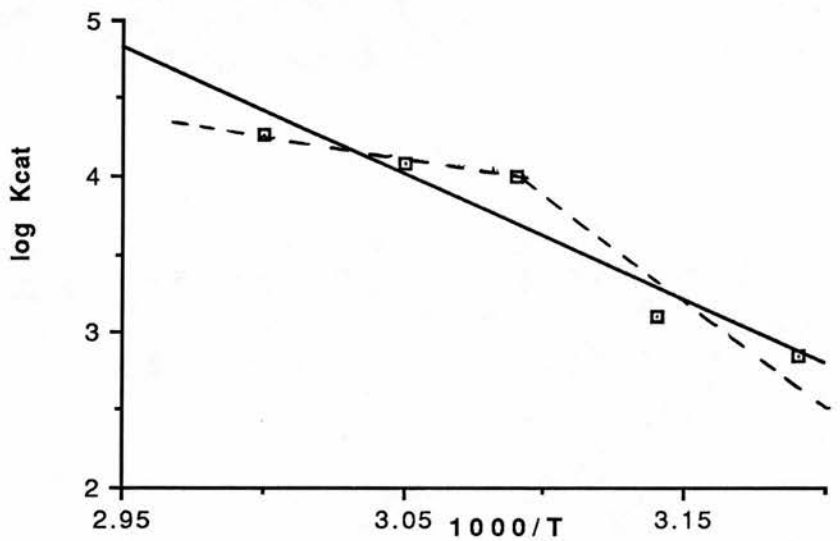


Figure 4.7: Arrhenius plot of temperature dependence of K_{cat} for the *T. acidophilum* pyruvate kinase. The activity of the enzyme was determined using the direct spectrophotometric assay of Pon and Bondar (1967). The pH of the assay buffer was adjusted to 7.5 at each temperature. K_{cat} values were calculated according to Cornish-Bowden (1979). The dotted line shows the possible 'kink' in the plot (see text for discussion).

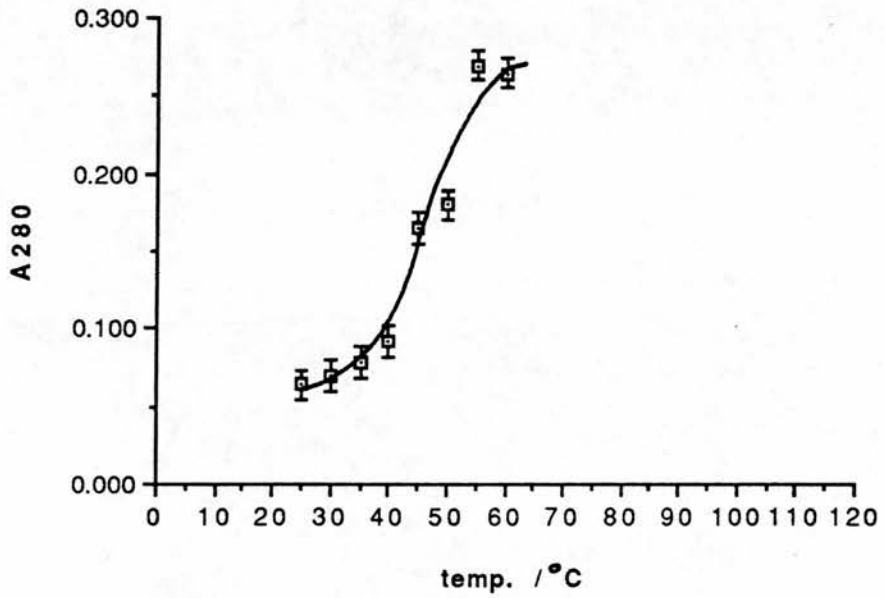


Figure 4.8: UV absorbance at 280 nm of the *T. acidophilum* pyruvate kinase as a function of temperature.

which two activation energies can be calculated. The 'kink' in the plot between the two states of the enzyme appears to occur at a temperature of 324 K, or 51°C (illustrated by the broken line in Fig. 4.7) although better data would be required to be certain of this.

The plot of optical density at 280 nm of a solution of the purified pyruvate kinase, shown in Fig. 4.8, reflects the change in solvation of key chromophores within the protein (such as the exposure of Tyr and Trp residues to solvent water). The dramatic change in A_{280} for the enzyme which occurs at around 50°C supports the observation derived from the Arrhenius plot that a conformational change may be displayed by the pyruvate kinase at that temperature. The implications of all of these observations are discussed later.

4.3.2.3 Regulation of the pyruvate kinase

Mesophilic pyruvate kinases are known to be subject to fine control by a range of effectors, some of which act allosterically. Typical allosteric regulators include AMP (*Bacillus* enzyme, Sakai *et al.*, 1986) and fructose-1,6-bisphosphate (yeast enzyme, Murcott *et al.*, 1991) Initial experiments suggested that the *T. acidophilum* pyruvate kinase was unaffected by such regulators. It was subsequently discovered, however, that the commercial preparation of ADP being used (Sigma) contained contaminating amounts of AMP which could be sufficient to activate the enzyme. Hence, prior to kinetic constant determinations it was necessary to purify the Sigma ADP using the method of Hunger and Reinbothe (1974, described in 2.2.9). The elution profile of the Pharmacia K9/60 column containing the cation exchanger DEAE-Sephadex A-25 used for the purification is shown in Fig. 4.9. Elution of the column was followed by % transmittance measurements at 254 nm. 10 mg of purified ADP eluted from the column in a volume of 100 ml (10 x 10 ml fractions). 1 ml was then removed from the 10 ml fraction containing the highest concentration of pure ADP and dried down in a vacuum dessicator to 100 µl. This contained the 0.1 mg of ADP required for the pyruvate kinase assay procedure.

To investigate possible allosteric regulation of the *T. acidophilum* pyruvate kinase by adenosine phosphates, kinetic constants were then determined for the enzyme under three sets of conditions; first using

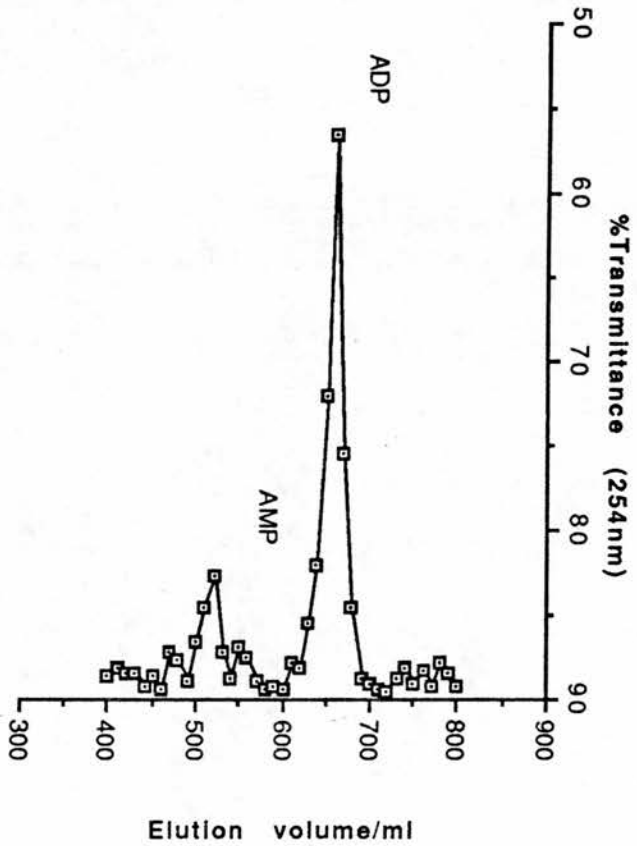


Figure 4.9: Separation of 5'-nucleotides on a DEAE-Sephadex A-25 column. The matrix was packed into a Pharmacia K9/60 column (0.9 x 60 cm). 10 mg of Sigma ADP was added to the column and was eluted using a linear gradient of 0-0.3 M ammonium carbamate. The flow rate of the column was 42 ml/hr and the elution was followed by transmittance at 254 nm (Hunger and Reinbothe, 1974).

only pure saturating ADP, secondly using ADP + 30 μM AMP (i.e. unpurified Sigma ADP) and thirdly using ADP + 10 mM ATP. The resultant saturation curves, obtained using PEP as variable substrate, are shown in Fig. 4.10. Using pure ADP and ADP+ATP, half-saturation PEP concentrations of around 0.7 mM were obtained. Using unpurified commercial ADP, which contains 30 μM AMP, a hyperbolic curve is seen, which suggests that AMP is acting as an allosteric activator of the pyruvate kinase. The K_m value for this curve, estimated from the Hanes plot shown in Fig. 4.11, is 0.043 mM. This latter observation was confirmed using ADP as the variable substrate (with PEP saturating) both in the presence and absence of AMP (Fig. 4.12a) . From these saturation curves, the K_m values for the enzyme using ADP as substrate with or without AMP were 0.1 mM and 0.68 mM respectively. The hyperbolic shape of the curve is confirmed by the linear plot of the data shown in Fig. 4.12b. The Hanes plots for -AMP (and +ATP) varying PEP did not show simple linear relationships (Figs. 4.13 and 4.14) and therefore, to establish whether the enzyme shows cooperativity, the Hill coefficient (h) of the enzyme was determined. A Hill plot (log(v/Vmax-v) vs. log[PEP]) of the data from the PEP saturation curves was constructed. The theoretical Vmax (0.041 absorbance units/min) used was estimated from Hanes plots. Three Hill plots are shown (Figs. 4.15, 4.16 and 4.17) for data obtained using PEP as variable substrate in the presence of ADP alone, ADP + AMP activator and for ADP + ATP. The Hill coefficients, estimated from the straight line portions of the plots (around the 50% saturation point), were all approximately 1.0. An enzyme showing no cooperativity towards its substrates is expected to have a Hill coefficient of 1.0. These results, therefore, indicate that whilst the pyruvate kinase appears to be activated by AMP, there is no evidence for positive cooperativity towards it substrates nor for allosteric regulation of the enzyme *in vitro*. Whether the pyruvate kinase is subject to such regulation *in vivo* depends on whether the concentrations of the putative regulators change over the range found in the cytoplasm of *T. acidophilum*.

4.3.2.4 Thermostability of the pyruvate kinase

As an adjunct to temperature profile experiments, thermostability studies were also conducted in which the enzyme was incubated at specified temperatures for up to 30 min. prior to activity assays. This was

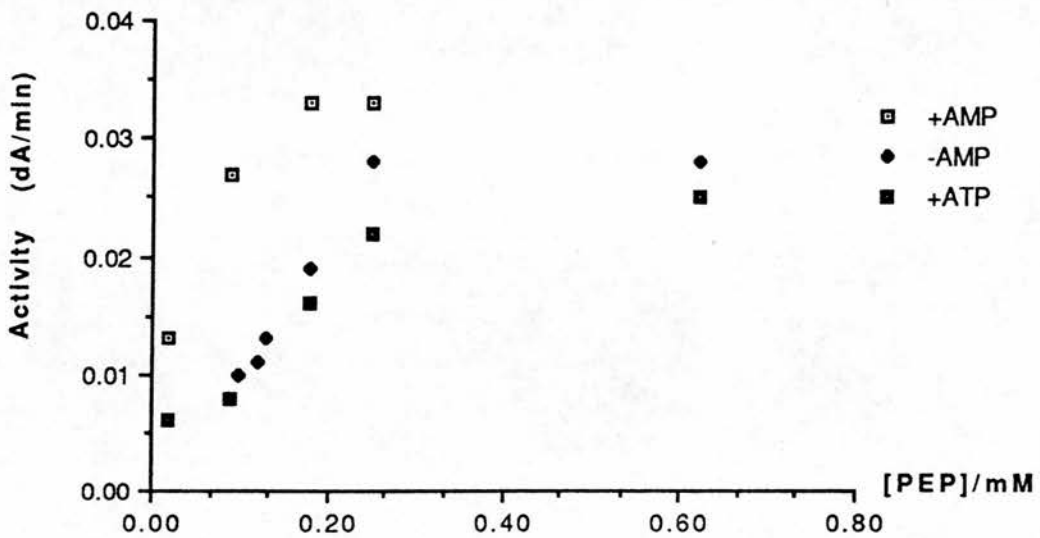


Figure 4.10: Saturation curves for the *T. acidophilum* pyruvate kinase. Dependence of activity on [PEP] was monitored in the presence and absence of allosteric effectors. The activity of the enzyme was determined at 60°C using the direct assay of Pon and Bondar (1967).

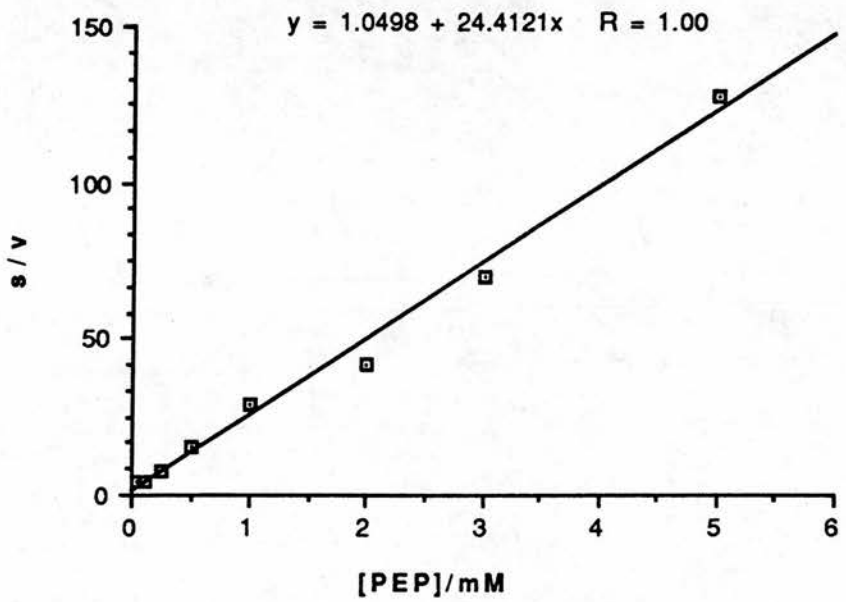
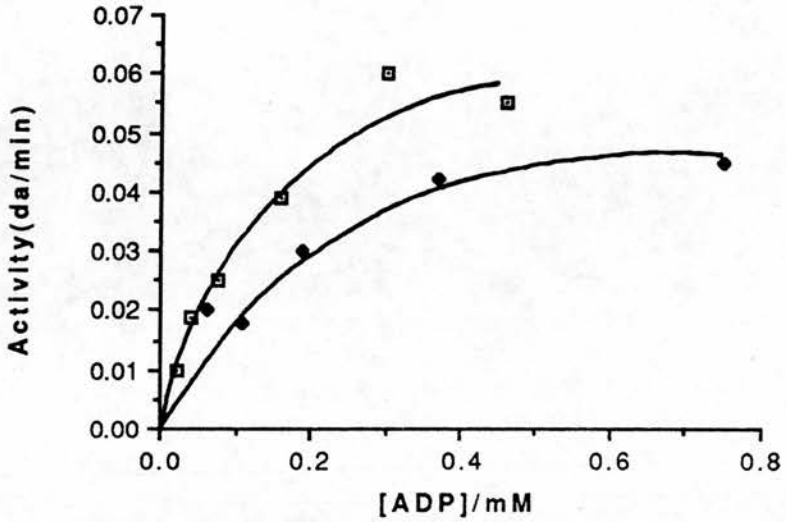
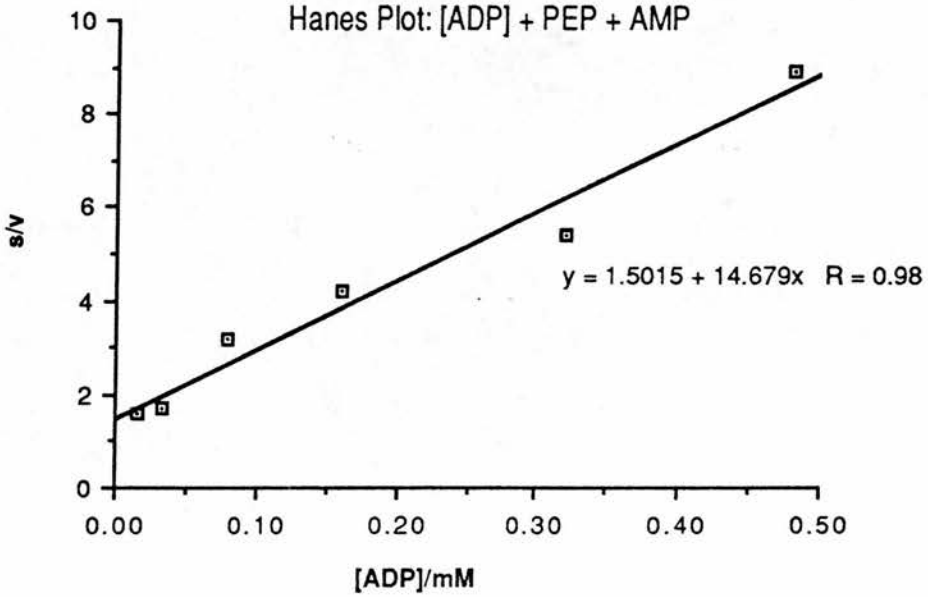


Figure 4.11: Hanes plot for the *T. acidophilum* pyruvate kinase using PEP as variable substrate in the presence of 30 μM AMP.

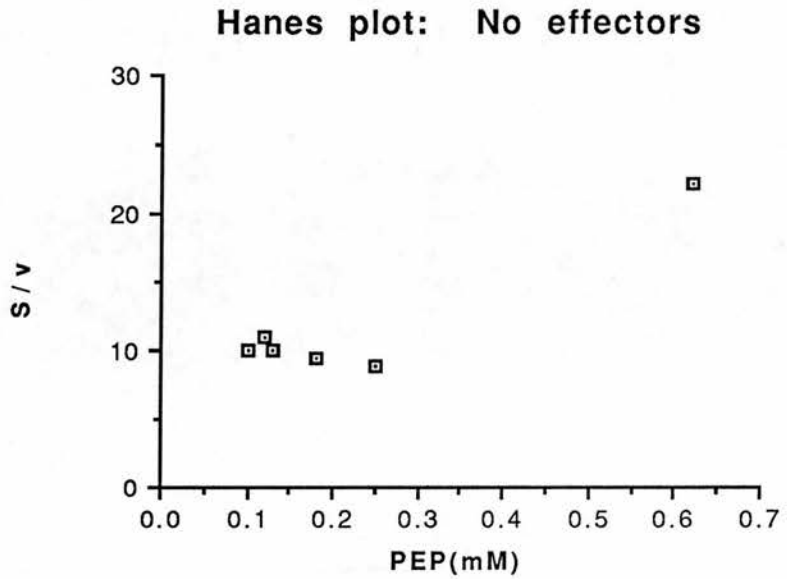
Figure 4.12a. Saturation curves for the *T. acidophilum* pyruvate kinase. Dependence of activity on [ADP] was monitored in the presence and absence of 30 μM AMP. The enzyme activity was determined at 60°C using the direct assay of Pon and Bondar (1967). Squares: +AMP; diamonds: no effectors.

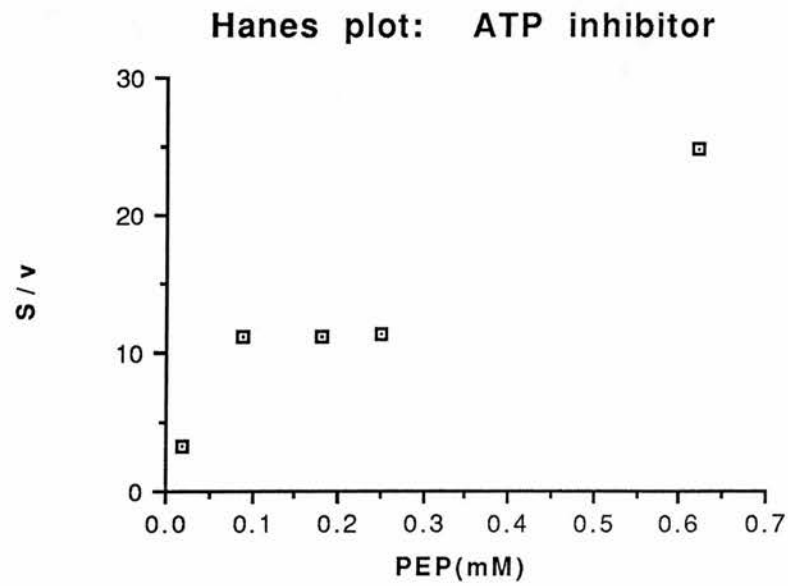
Figure 4.12b. Hanes plot for variable ADP in presence of AMP.

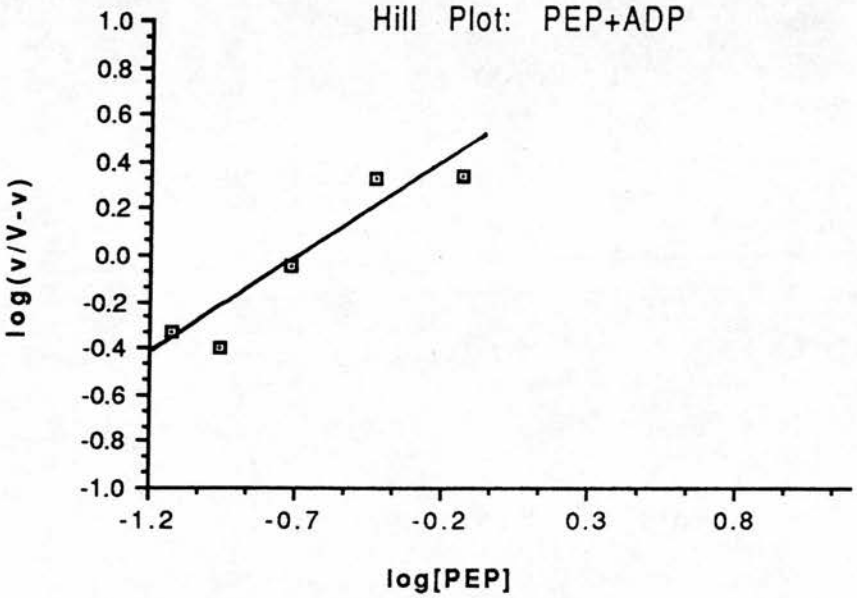

Activity(da/min)
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0.0
0.2
0.4
0.6
0.8
[ADP]/mM


Hanes Plot: [ADP] + PEP + AMP
s/v
10
8
6
4
2
0
0.00
0.10
0.20
0.30
0.40
0.50
[ADP]/mM
y = 1.5015 + 14.679x R = 0.98

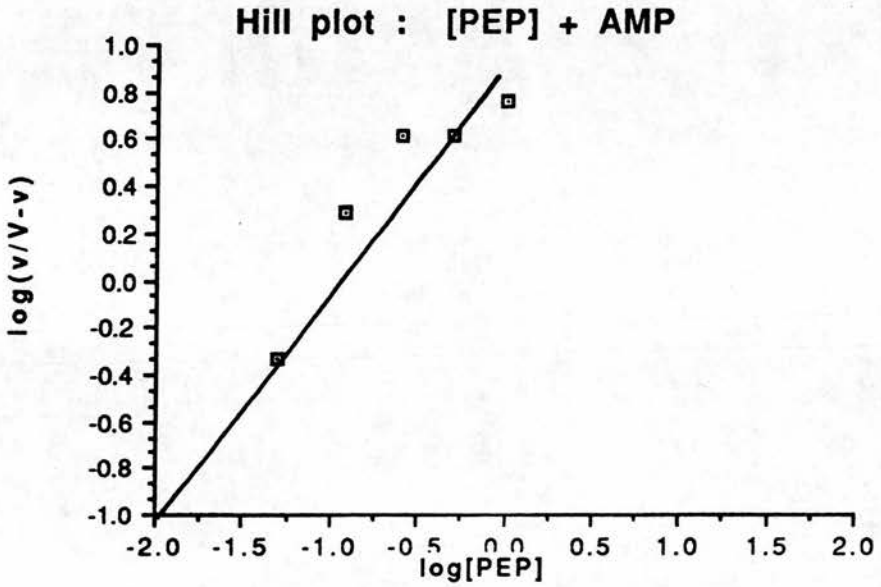
Figures 4.13 and 4.14: Hanes plots for variable PEP in absence of effectors and in presence of ATP respectively.

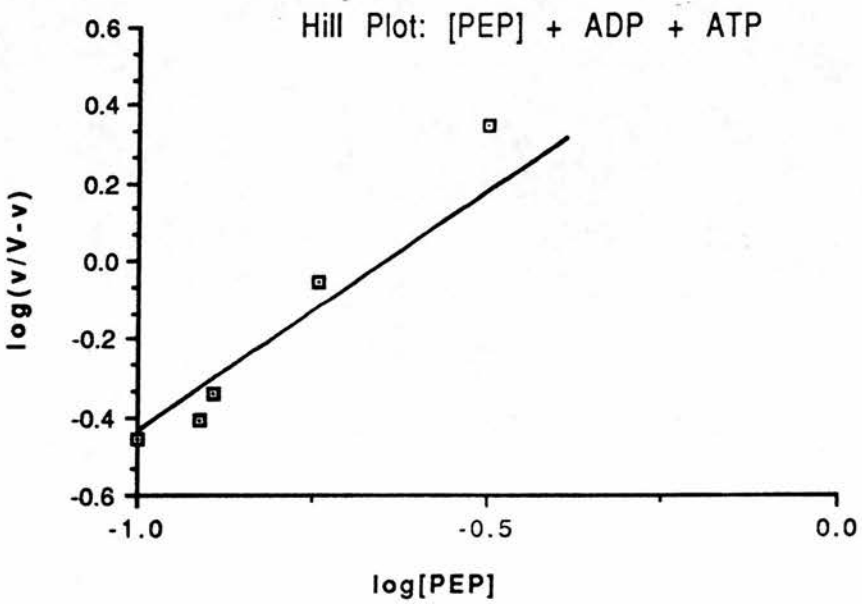

Hanes plot: No effectors
S / v
30
20
10
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
PEP(mM)


Hanes plot: ATP inhibitor
S / v
30
20
10
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
PEP(mM)



Figures 4.15-4.17: Hill plots for the *T. acidophilum* pyruvate kinase. In each case PEP was used as variable substrate. *Figure 4.15* is the curve obtained for the enzyme in the absence of effector molecules; *figure 4.16* is the curve obtained using 30 μM AMP; and *figure 4.17* is the curve obtained using 10 mM ATP. The lines shown have a gradient of 1.0 in keeping with the observation that the enzyme shows no apparent cooperativity to its substrates.


Hill plot : [PEP] + AMP
log(v/V-v)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
log[PEP]


Hill Plot: [PEP] + ADP + ATP
log(v/V-v)
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-1.0
-0.5
0.0
log[PEP]

to determine the resistance of the enzyme to thermal denaturation. Fig. 4.18 shows the stability of the pyruvate kinase at three different temperatures: 60°C, 65°C and 70°C. For comparison, rabbit muscle pyruvate kinase (Sigma) was also tested for stability at 60°C. At 60°C, the growth temperature of *T. acidophilum*, the enzyme from that organism was found to be very stable, resisting a 30 min. incubation without appreciable loss of activity. After 30 min. at the same temperature, the rabbit muscle enzyme has lost approximately 94% of its activity. At the higher temperatures, the *T. acidophilum* pyruvate kinase does lose activity gradually. After 30 min. at 65°C, the enzyme retains only 60-70% of its activity whereas at 70°C only 15% of the activity remains after 20 min. and it is completely abolished after 30 min.

A summary table of the results presented above is shown in Table 4.2, with mesophilic enzyme data for comparison - see discussion section 4.5.

4.4 Discussion 1: Purification

From Table 4.1, it can be seen that a substantial purification (500 fold) of the enzyme has been achieved. At least 10 runs of the purification procedure were performed and, on average, the final yield of pure pyruvate kinase from 1l of stationary phase culture (200 mg of *T. acidophilum* cells) was approximately 50 µg (830 pmol). From SDS-polyacrylamide gels of crude cell extract, it was estimated that the enzyme comprises approximately 2% of the total cellular protein, a figure typical for glycolytic enzymes isolated from many organisms (Fothergill-Gilmore, 1989). The exact percentage yield of the protein was impossible to estimate, however, since it appears that a compound released in the crude cell extract (possibly free ATP) inhibits the enzyme; this is evidenced by an increase in total enzymic activity at the 60% (w/v) ammonium sulphate precipitation stage (see Table 4.1). The final recovery of activity after this step is 75%, however this figure is almost certainly also an overestimate as ammonium sulphate is known to inhibit many enzymes, including pyruvate kinases.

The specific activity of the final, homogeneous preparation of the enzyme was 200 µmol NADH converted min-1 (mg of protein)-1. This is rather low in comparison with pyruvate kinases isolated from eukaryotic

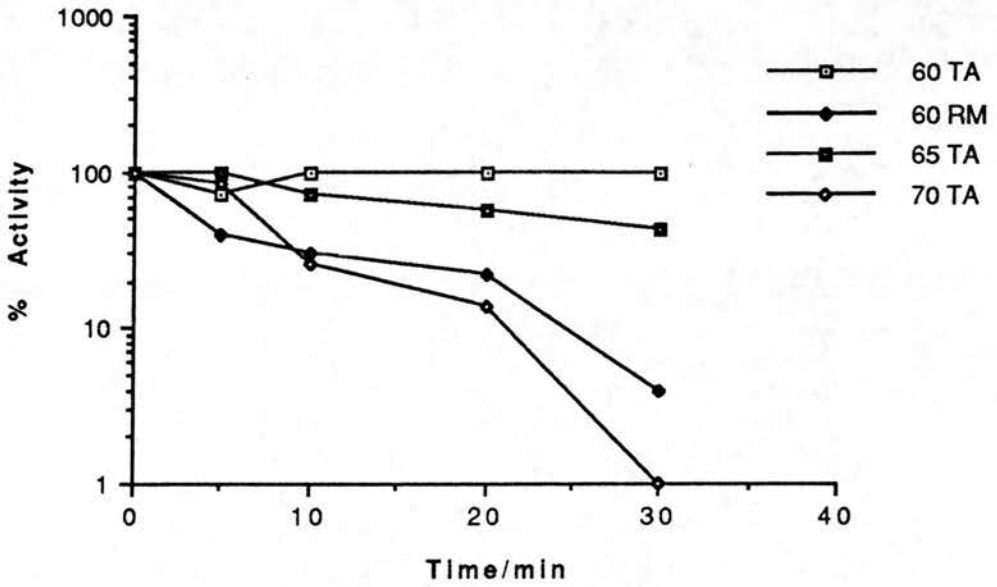


Figure 4.18: Thermostability curves for the *T. acidophilum* pyruvate kinase. The enzyme from *T. acidophilum* was incubated at 60, 65 and 70°C and the activity of the enzyme was determined after sampling at regular time intervals (in each case the assay of Bucher and Pfleiderer was used at 55°C). For comparison, the pyruvate kinase from rabbit muscle (Sigma) was incubated at 60°C and its activity was determined at the same time intervals (curve marked 60 RM).

property	T.acido	B.stearo.	Yeast
Specific Activity (U/mg)	**201**	333	367
Temperature (^{o}C)	**60**	60	37
of assay			
$S_{0.5}$ [PEP] (mM)	**0.64**	2.0	3.7
K_m [PEP] +AMP (mM)	**0.043**	0.2	0.16
$S_{0.5}$ [PEP] +ATP (mM)	**0.70**	-	-
K_m [ADP] +AMP (mM)	**0.1**	-	0.54
$S_{0.5}$ [ADP] + ATP (mM)	**0.68**	-	-
Hill Coefficient (PEP)	**~ 1.0**	1.8	2.86
pH Optimum (-AMP)	**7.5**	7.2	6.0-6.5
Subunit Mr x10^{-3}	**60**	62	59
Blocked N-terminus?	**yes**	yes	yes

Table 4.2: Physical and kinetic properties of the pyruvate kinase

Units of enzyme activity μmol/min. '-' indicates value not measured. $S_{0.5}$ is the dissociation constant for the enzyme-substrate complex. K_m is not used as this would imply typical Michaelis-Menten kinetics. B.stearo., *Bacillus stearothermophilus*, T.acido.,*Thermoplasma acidophilum*. Values for the *Bacillus* enzyme were from Sakai *et al.* (1986) and those for the yeast (*Saccharomyces cerevisiae*) enzyme from Murcott *et al* (1991).

(*Saccharomyces cerevisiae* - 367 U/mg, Murcott *et al.*, 1991) and bacterial (*Bacillus stearothermophilus* - 333 U/mg, Sakai *et al.*, 1986) sources. There is no obvious explanation for this difference. Possibly, some catalytic efficiency may have been sacrificed for the increase in structural rigidity which is often found in thermostable enzymes and particularly archaebacterial ones (Brock, 1967; Fontana, 1988; Vihinen, 1987): the activity of the enzyme may, therefore, be the result of an evolutionary trade-off between stability at high temperature and catalytic efficiency. This is unlikely, however, as whilst thermophilic enzymes are indeed more rigid than their mesophilic counterparts at low temperatures, they acquire the flexibility which is consistent with an optimum catalytic efficiency at high temperatures. A more likely explanation for the comparatively poor specific activity of the *T. acidophilum* pyruvate kinase is that it reflects the antiquity of the enzyme and the organism from which it has been obtained. That is to say, the pyruvate kinase of the archaebacterium has not evolved to the same efficient extent as have its mesophilic counterparts which suggests that the pyruvate kinase of *T. acidophilum* has not been subjected to strict selection pressure over the millenia.

4.5 Discussion 2: Characterisation

The work detailed in this chapter suggests that *T. acidophilum* pyruvate kinase appears to closely resemble its mesophilic counterparts, at least in terms of its physical and kinetic properties (Table 4.2). Its subunits each have an Mr of 60k. Like the bacterial (*B. stearothermophilus*) enzyme, it is activated by AMP and does not appear to be affected by fructose-1,6-bisphosphate (unlike the eukaryotic enzyme - *Saccharomyces cerevisiae*). This latter observation is not entirely surprising, however, given the non-phosphorylated glycolytic pathway of which the enzyme forms a part: fructose-1,6-bisphosphate is not an intermediate in the early part of the pathway used by *T. acidophilum* and other thermophilic archaebacteria (Danson, 1988). This is in contrast with the eukaryotic case, where fructose-1,6-bisphosphate is an intermediate of the Embden-Meyerhof pathway and is involved in feed-forward activation of pyruvate kinase.

AMP activates the enzyme (30 μM as compared to 140 μM AMP required for full activation by the *B. stearothermophilus* enzyme). This

activation can be explained either by the concerted model of Monod, Wyman and Changeux (MWC, 1965) or by the sequential model of Koshland, Nemethy and Filmer (KNF, 1966).

In the MWC model, the binding of a molecule of AMP to one of the subunits of the tetramer will alter the equilibrium between the tense (T, unligated, low affinity) and relaxed (R, ligated, high affinity) states of the enzyme in favour of the relaxed state. In order to simplify the mathematics of the concept, the quaternary structure of the enzyme is assumed to be always symmetrical and hence the enzyme is never in a state in which one of its subunits is in the R state and the others are in the T state. The binding of AMP to one of the subunits of the pyruvate kinase, therefore, puts the entire tetramer into the high affinity state and facilitates the binding of substrates to the entire molecule.

In the KNF model, the progress from the T state to the R state is assumed to be sequential and the assumption of symmetry is thus avoided. In this case, the allosteric activation of the pyruvate kinase can be explained as follows: AMP binds to one of the subunits of the enzyme causing a conformational change. This change is passed to adjacent subunits via intersubunit contacts and increases the affinity of those subunits for the substrates of the catalysed reaction. It is worth noting here that both models are simplifications and that the real processes involved in the allosteric activation of the pyruvate kinase are certainly much more complex than either suggests (Eigen, 1967). The enzyme has a Hill coefficient of approximately 1 which shows that it does not exhibit positive cooperativity towards its substrates.

In two respects, the *T. acidophilum* pyruvate kinase does show unusual properties, both of which are related to its enhanced temperature stability. First, it is the most thermostable pyruvate kinase yet isolated, being capable of catalysing conversion of PEP to pyruvate at temperatures of up to 90°C (Fig. 4.6). Secondly, the Arrhenius plot of the dependence of k_{cat} for the enzyme on the temperature may be kinked (Fig. 4.7). This type of plot is particularly difficult to interpret as, under the assay conditions used, the Km for the reaction (a complex of forward and back rate constants) can also vary with temperature because all of its component k values will. Furthermore, since the equilibrium constant of each step of the reaction may vary with temperature, it cannot even be

assumed that the forward and backward rate constants for each step will share the same temperature dependence.

Given the likely experimental errors, therefore, the data in Fig. 4.7 may be adequately represented by a straight line as shown - the expected result for an enzyme. It is worth noting here, however, that a kinked Arrhenius plot is not an unusual result, particularly for data obtained using thermostable enzymes (Ragone *et al.*, 1992; Arnone *et al.*, 1992). This phenomenon is thought to reflect a conformational transition in the thermophilic enzyme at the temperature at which the 'kink' in the curve occurs (note: other explanations for the 'kink' do exist - a change in the rate-limiting step; the existence of the enzyme as a mixture of isozymes; an effect of temperature on the substrates). The Arrhenius plot for the pyruvate kinase appears to 'kink' at about 50°C suggesting that a conformational change may occur at this temperature. This translates as a decrease in activation energy for the enzyme at temperatures above 50°C from 228 $kJmol^{-1}$ to 53 $kJmol^{-1}$ and corresponds well with the observation that enzyme activity is very low below that temperature. Further evidence for a temperature dependent conformational transition is presented in Fig. 4.8. The possible heat activation of the enzyme at around 50°C is accompanied by a rise in the absorbance of the protein at 280 nm. This suggests that a conformational shift in the structure of the pyruvate kinase is occurring at this temperature, exposing more chromophoric residues to the solvent. This may be necessary to extend the intially inactive conformation of the protein at lower temperatures to gain enzymic activity. Evidence of this type has recently been employed to explain the heat activation of another thermophilic enzyme from an archaebacterium - glutamate dehydrogenase from the hyperthermophile *Pyrococcus furiosus* (Klump *et al.*, 1992).

Such conformational transitions are brought about by a number of factors. For example, numerous physico-chemical properties of amino acids which are intimately involved in determining protein architecture (Creighton, 1984), are temperature dependent. The hydrophobicity of each amino acid does not change linearly with temperature and differences in the thermal expansion coefficient of the protein matrix may also lead to internal structural tensions. Amongst others, these may represent forces which are able to affect structural reorganisation of the pyruvate kinase.

Speculating further, the conformational change from an active to an inactive form for certain thermophilic enzymes may represent a temperature dependent 'molecular switch' of the type postulated for the aspartate amino transferase of *Sulfolobus solfataricus* (Facchiano *et al.*, 1992). *T. acidophilum* exists naturally in acidic hot puddles in which zones of different temperature are likely to exist. If the temperature of a cell's environment decreases from 50°C to 40°C, such a 'molecular switch' might regulate the activity of the pyruvate kinase and, therefore, the glycolytic metabolic rate of the organism, in such a way as to leave it in a latent form. The possible advantages of such a switch would be uncertain, other than the fact that many organisms, when faced with unfavourable environmental conditions, resort to a form of 'hibernation' to wait in a low metabolic rate state until conditions improve. It is worth noting that the conformational change interpretation of the Arrhenius and temperature profile data postulated here would not constitute the only example of such a phenomenon amongst the pyruvate kinases. The enzyme purified from the Alaskan king crab (Somero and Hochachka, 1968) is also known to exist in two forms. The 'cold' variant exhibits hyperbolic kinetics with PEP and has a minimum K_m at 5°C. The warm variant displays sigmoidal kinetics and has a minimum K_m at 12°C. It is important to stress that the above discussion is speculation based on extremely preliminary work and that much more evidence, particularly structural, is required before any of these ideas can be tested.

4.6 Summary of conclusions

The main conclusions from this work are twofold. First, the pure pyruvate kinase of *T. acidophilum* is kinetically and physically closely related to its mesophilic counterparts, especially those isolated from bacterial sources. The enzyme displays comparable kinetic constants (when assayed at 60°C, the growth temperature of the archaebacterium) to those found with mesophilic pyruvate kinases at physiological temperature. Secondly, in spite of these similarities, the enzyme does exhibit some unusual properties. It is highly stable at elevated temperatures (Fig. 4.6 suggests that the temperature optimum for the enzyme may be anywhere up to 90°C, although the actual concept of an 'optimum' temperature is considered by many workers to be a

meaningless one) and, correspondingly, it may be structurally more rigid than its mesophilic counterparts.

Finally, there is tentative preliminary evidence to suggest that the enzyme may exist in two states, an active one above 50°C and a relatively inactive one below that temperature. Much better evidence is obviously required before any concrete conclusions can be drawn from this work and hence an investigation, using structural studies such as circular dichroism or fluorescence quenching at different temperatures for example, would be an interesting next step.

Chapter 5: Structural Characterisation of the *T. acidophilum* Pyruvate Kinase

CHAPTER 5: STRUCTURAL CHARACTERISATION OF THE *T. ACIDOPHILUM* PYRUVATE KINASE

5.1 Introduction

Following the preliminary physical and kinetic characterisation of the purified *T. acidophilum* pyruvate kinase (described in the previous chapter), the next step towards full characterisation of the protein was the determination of certain primary structural features. The amino acid composition of the protein was determined, first to assist in the choice of proteolytic enzymes for the subsequent digestion of the pyruvate kinase into peptides for microsequence analysis and secondly, for comparison with the amino acid compositions of mesophilic pyruvate kinases. Partial amino acid sequence from the protein was also determined for use in the later cloning and nucleotide sequencing of the corresponding stretch of *T. acidophilum* genomic DNA.

The partial amino acid sequence obtained was to be used in two ways. First, the sequence was used, in combination with codon usage tables for other *T. acidophilum* protein encoding genes, to design oligonucleotide probes (and PCR primers) in order to identify the gene for the pyruvate kinase in restriction digests ('mini-libraries') of *T. acidophilum* genomic DNA. Secondly, knowledge of peptide sequences derived from different parts of the protein was to be used as a check during the DNA sequencing for the occurrence of any frame-shifting errors. It was considered that this joint approach to the sequencing of the pyruvate kinase, comprising a combination of partial peptide sequencing and full nucleotide sequencing, would be a more effective and accurate route to complete sequence analysis than either technique alone (the explicit cloning strategy used is described in Chapter 6).

This chapter is thus divided into sections describing amino acid analysis, preliminary sequencing attempts, cleavage of the pyruvate kinase and separation and sequencing of the resultant peptides.

5.2 Results

5.2.1 Amino acid analysis

An accurate amino acid analysis of a protein is important as it

provides a yardstick against which the sequence analysis of the whole polypeptide chain can be compared i.e. it provides a check for nucleotide sequencing of the coding gene for the protein, predicting the numbers of each amino acid residue which should be found in the complete primary sequence. Furthermore, it allows prediction, prior to sequencing, of the best method for digestion of the protein into peptides.

The determination of the amino acid composition of the pyruvate kinase involved two stages: quantitative hydrolysis into constituent amino acids using 6 M HCl (a process which destroys tryptophan), and subsequent determination of the released amino acids. The second stage involved preliminary conversion of the free amino acids to phenylthiocarbamyl (PTC) derivatives with the reagent phenylisothiocyanate (PITC). PITC reacts with the free amine groups released by hydrolysis of the protein to yield PTC-amino acids which absorb strongly in the ultraviolet region. This step is necessary prior to the separation and detection of the amino acid residues as most amino acids do not possess significant UV absorbance or fluorescence at wavelengths suitable for detection.

Many compounds, including impurities in solvents and by-products of the reaction also absorb in the same region of the UV spectrum as the PTC-derivatives and hence contamination of the sample was avoided with great care. To facilitate this, the homogeneous pyruvate kinase preparation from the affinity column was further purified on an Aquapore RP-300 (7 μm particle size; 2.1 mm x 30 mm) column. 100 pmol of the enzyme was loaded onto the column which was then eluted with a linear 8-80% (v/v) gradient of acetonitrile in aqueous 0.1% (v/v) TFA. Elution of the protein was followed at 220 nm and the peak was collected directly from the column in a microcentrifuge tube in as small a volume as possible (approximately 30 μl). The elution profile of the pure pyruvate kinase from the column is shown in Fig. 5.1. The full 100 pmol was then hydrolysed and 50 pmol was subjected to derivatization and chromatographic separation. To avoid contamination of samples during handling, the last three steps of the procedure were performed in a fully automated instrument, the Applied Biosystems 420A Derivatizer, which was calibrated previously with a known amount (50 pmol) of each amino acid derivative.

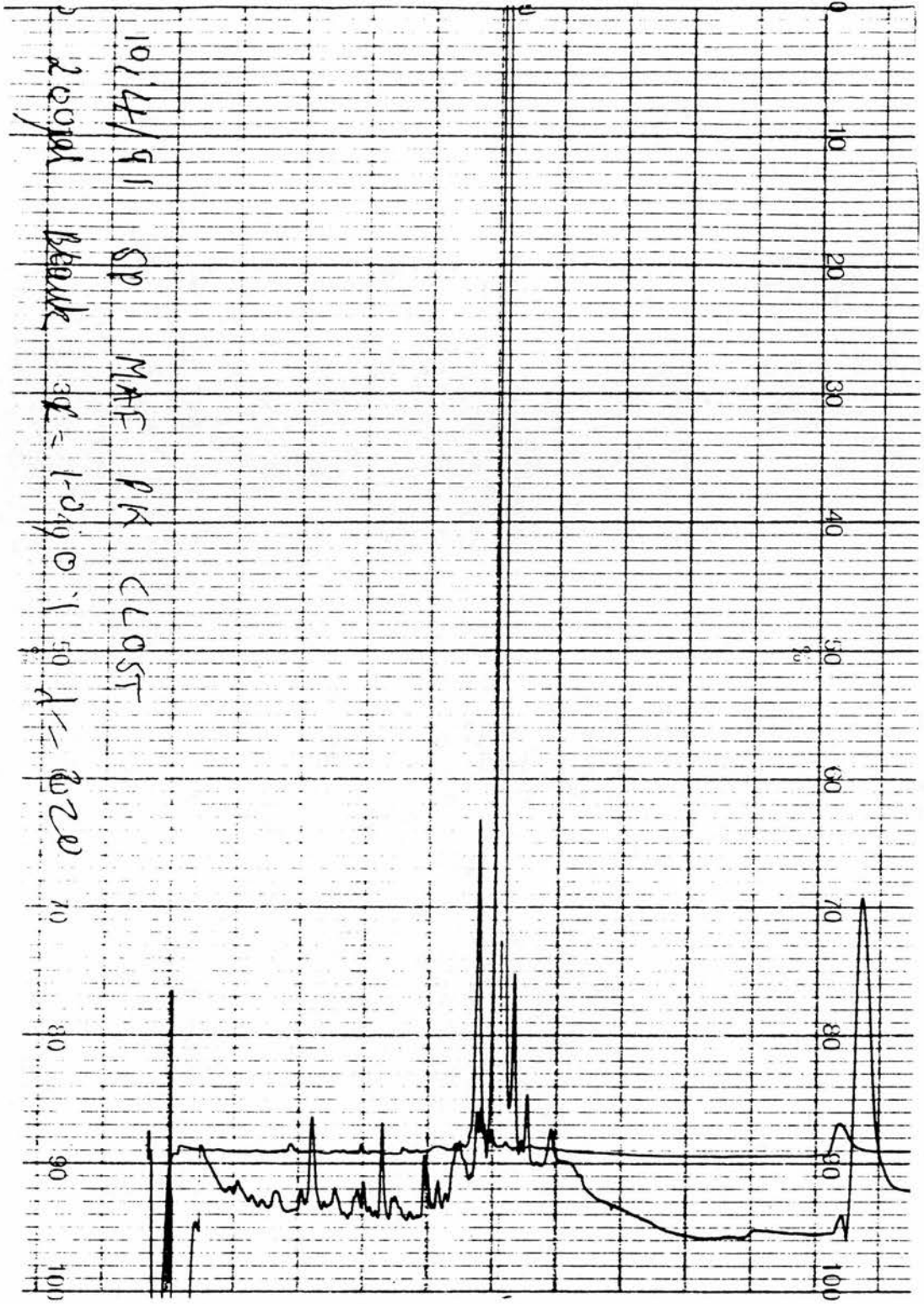


Figure 5.1: Elution of purified *T. acidophilum* pyruvate kinase from the Aquapore RP-300 column.

The chromatograph shown in Fig. 5.2 is a typical separation profile of the standard mix of amino acids detected at 254 nm. Since each amino acid residue is identified by its retention time on the column in the Derivatizer, it is very important to do a separate calibration run for each batch of samples as these times can change slightly between determinations. Table 5.1 is an interpretation of the results from the standard separation and gives a list of retention times and the corresponding peak areas obtained for 50 pmol of each PTC-amino acid derivative. Table 5.2 is the chromatograph report for 50 pmol of the purified *T. acidophilum* pyruvate kinase. The table gives the peak area values for each PTC-amino acid derivative, the corresponding pmol amounts of each residue (corrected using the internal calibration standard) and the calculated mol% of each residue in the protein (obtained by determining the percentage of the total pmol value of the pmol amounts of each amino acid). Table 5.3 shows these mol% values and their corresponding actual amino acid numbers for the *T. acidophilum* pyruvate kinase (actual numbers were based on assuming a total number of amino acids for the enzyme of approximately 580, from the known Mr of 60k, and then converting the mol% values into numbers of amino acids, rounding them off to the integer). For comparison, both mol% values and amino acid numbers for pyruvate kinases from Bacillus stearothermophilus and chicken muscle are also shown. These results are analysed further and enlarged upon in the later discussion section.

5.2.2 Preliminary sequencing attempts

A usual first step in the elucidation of the amino acid sequence of a protein is the determination of its amino-terminal residue. Various methods for this exist including the dansyl chloride method and the use of PITC incorporation. The latter method was used for these studies, given the departmental access to an Applied Biosystems 477A instrument (which was coupled to a 120A phenylthiohydantoin analyser in the WELMET protein sequencing facility). 200 pmol of the *T. acidophilum* pyruvate kinase eluted from the 5'AMP affinity column was purified further, as for amino acid analysis, on an Aquapore RP-300 column. The pure protein was then treated as described in section 2.2.13 and 2.2.14 prior to Edman degradation (this is described in section 5.2.5). No sequence from the protein was obtained. The most probable explanation for this is that many proteins, including several pyruvate kinases, have blocked N-terminal residues. This phenomenon is commonly due to acetylation or cyclization

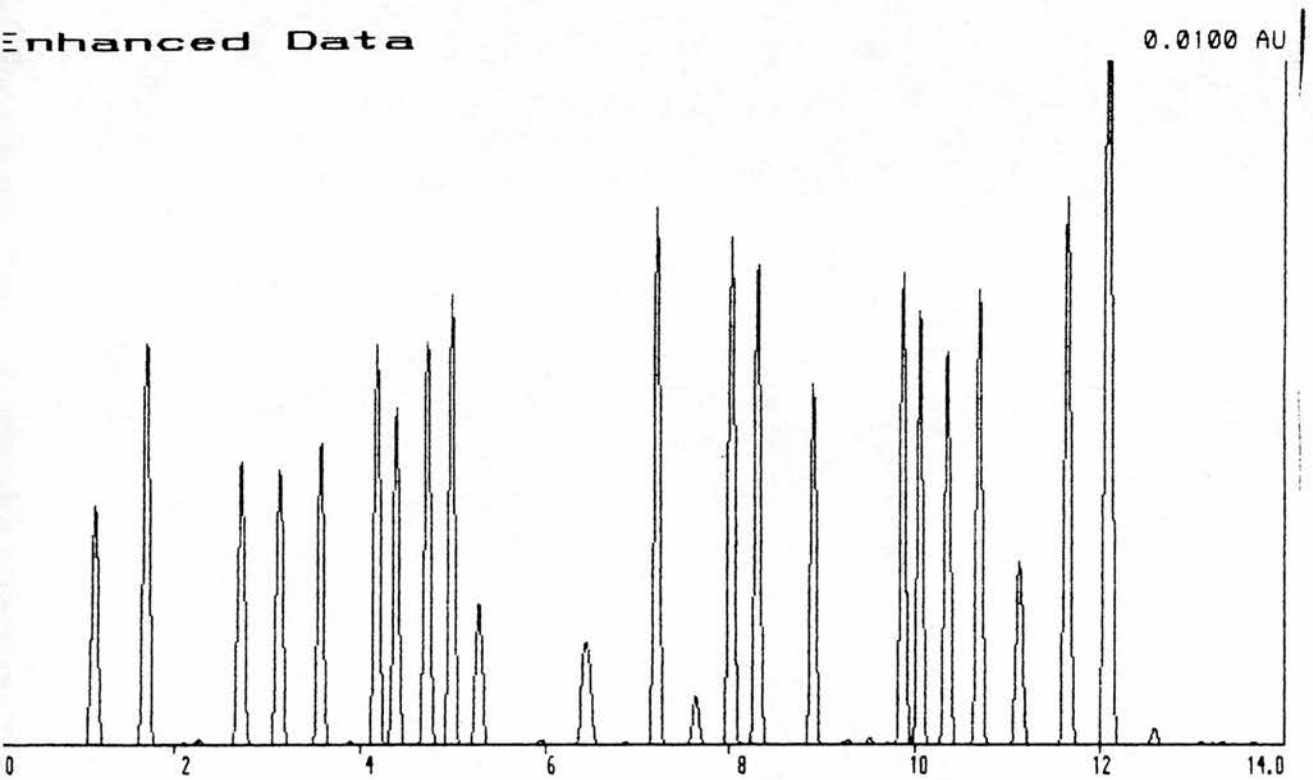


Figure 5.2: Calibration chromatogram for the Applied Biosystems 420A amino acid analyser. Each peak detected at 254 nm corresponds to 50 pmol of a standard amino acid derivative. The order of amino acids eluted is: D, E, S, G, H, R, T, A, P, Y, V, M, C, I, L, F, K.

amino acid	retention time	peak area	pmol by area
Asx	1.16	68226	50.00
Glx	1.70	102064	50.00
Ser	2.72	78568	50.00
Gly	3.13	77126	50.00
His	3.57	78734	50.00
Arg	4.17	101058	50.00
Thr	4.38	85250	50.00
Ala	4.72	104750	50.00
Pro	4.97	114326	50.00
Tyr	7.21	133664	50.00
Val	8.02	126388	50.00
Met	8.29	120588	50.00
Cys	8.90	88770	50.00
Ile	10.06	104360	50.00
Leu	10.35	99940	50.00
Phe	10.69	115594	50.00
Lys	11.65	144258	50.00

Table 5.1: Calibration of the amino acid analyser using a standard mix of 50 pmol of amino acid derivatives. Peak areas are shown in uAU. Retention times are shown in min. Minimum peak threshold was 1000 uAU.

amino acid	pmol corrected	pmol by area	peak area	mol%
asx	34.94	29.18	57028	7.16
glx	56.61	47.28	92403	11.60
ser	23.33	19.48	38071	4.78
gly	64.95	54.24	106005	13.31
his	11.66	9.74	19035	2.39
arg	17.47	14.59	28514	3.58
thr	29.13	24.33	47550	5.97
ala	48.31	40.35	78859	9.90
pro	31.62	26.41	51615	6.48
tyr	4.98	4.16	8130	1.02
val	42.46	35.46	69302	8.70
met	14.15	11.82	23100	2.90
cys	4.15	3.46	6762	0.85
ile	39.14	32.69	63888	8.02
leu	39.98	33.39	65256	8.19
phe	5.81	4.85	9478	1.19
lys	19.13	15.98	31231	3.92

Table 5.2: Amino acid composition of *T. acidophilum* pyruvate kinase. Peak areas are expressed in uAU. 100 pmol of purified pyruvate kinase was hydrolysed and then redissolved in 0.1% TFA. One half of the 80 µl was loaded into the analyser.

amino acid	chicken		*Bacillus*		*Thermoplasma*	
	mol%	residues	mol%	residues	mol%	residues
asx	9.45	50	9.59	56	7.16	42
glx	9.26	49	10.19	59	11.60	68
ser	4.73	25	5.25	31	4.78	28
gly	7.94	42	8.24	48	13.31	78
his	3.40	18	2.29	13	2.39	14
arg	6.05	32	4.62	27	3.58	21
thr	4.73	25	7.04	41	5.97	35
ala	10.96	58	10.74	63	9.90	58
pro	3.78	20	3.95	23	6.48	38
tyr	1.51	8	0.85	5	1.02	6
val	8.32	44	10.01	58	8.70	51
met	3.97	21	3.17	18	2.90	17
cys	1.51	8	0.29	2	0.85	5
ile	6.99	37	7.77	45	8.02	47
leu	6.80	36	8.62	50	8.19	48
phe	3.21	17	1.30	8	1.19	7
lys	6.80	36	5.89	34	3.92	23
TOTAL		529		586		582

Table 5.3: Amino acid composition of *T. acidophilum* pyruvate kinase and comparison with pyruvate kinases isolated from a eukaryotic and a eubacterial source. The mol% figures shown in table 5.2 were converted to actual amino acid numbers using the subunit Mr value of 60000 estimated from the SDS-polyacrylamide gel shown in Fig. 4.2(a) and the values rounded off to the integer.

of N-terminal glutaminyl residues to pyrrolidonecarboxylyl residues. Other common blocking groups include formyl-, trimethylalanyl- and fatty acyl-amino terminal groups. Blocking of this kind impedes the process of Edman degradation and as a result sequencing cannot proceed. The implications of this result are further discussed later.

5.2.3 Cleavage of the pyruvate kinase

The failure of the initial sequencing attempts to gain N-terminal protein sequence prompted the next step in the structural characterisation of the *T. acidophilum* pyruvate kinase: the selective cleavage of the protein into peptide fragments for sequencing. The amino acid analysis of the protein was used to indicate which cleavage methods were likely to be of most use. The content of approximately 6 mol% of each of the basic residues lysine and arginine in the protein suggested that either of the proteolytic enzymes trypsin or clostripain could be productively used. Equally, the presence in the pyruvate kinase of 17 methionine residues suggested that cyanogen bromide degradation would also be suitable.

Cleavage of the large pyruvate kinase subunits by both of these methods was predicted, from the amino acid analysis, to lead to complex mixtures of small peptides: proteolytic cleavage by either trypsin or clostripain would lead to a mixture of 22-24 peptides of average length 25 residues whereas chemical degradation by cyanogen bromide would lead to 18 peptides of average length 32 residues. As a consequence of this complexity of products, it was decided to resolve the cleavage peptides using reverse phase H.P.L.C. (or F.P.L.C.) chromatography prior to sequencing.

5.2.3.1 Cyanogen bromide cleavage

Since its introduction by Gross and Witkop (1961), the cleavage of polypeptides at methionine residues by cyanogen bromide (CNBr) under acidic conditions has been used frequently. Methionine residues are converted into a mixture of C-terminal homoserine residues and homoserine lactone residues which are interconvertible (Ambler, 1965). The reaction was performed as described in 2.2.18 for 24 hr at room temperature in 70% (v/v) formic acid in water. After completion of the reaction, the mixture was diluted with water and freeze-dried.

The peptide products of the reaction were redissolved in 0.1% (v/v) TFA and resolved on an Aquapore RP-300 column (attached to an Applied Biosystems 130A hplc machine) using a linear gradient of 8-80% acetonitrile in 0.1% (v/v) TFA. Peptide peaks were detected by absorbance at 210 nm (the peptide bond). The rationale of the separation is peptide resolution by hydrophobic interaction chromatography. The resultant peptide profile is shown in Fig. 5.3. Each of the peaks was collected from the column eluent in as small a volume as possible (typically <50 µl). Peak 'D' , considered to be the sharpest (and, therefore, presumably the cleanest) of the peaks, was selected for sequencing. The results of the sequencing experiment are described later.

5.2.3.2 Proteolytic (enzymatic) cleavage

Clostripain (EC 3.4.22.8) is a sulphydral protease which requires the presence of a thiol group for full activity. It is generally specific for arginine residues although it also shows some proteolytic activity towards lysine residues. To prevent this occurring, the lysine residues of the purified *T. acidophilum* pyruvate kinase were succinylated (see 2.2.17) prior to digestion with the clostripain. After digestion for 4 hr at 37°C in the presence of 1 mM dithiothreitol, using a weight ratio of enzyme to substrate of 1 : 50, the reaction was considered to be complete and the resultant peptides were separated using the H.P.L.C. system described above. The profile of the column eluent is shown in Fig. 5.4. In this particular case, the resolution of the peptides was poor and collection of a single peak proved to be impossible. To improve the separation of the peptides, the H.P.L.C. column step was replaced with an F.P.L.C. step using the PEP-RPC reverse phase column (Pharmacia). A prolonged linear gradient of 0-30% (v/v) acetonitrile in 0.1% (v/v) TFA was employed and the resultant improved separation of the peptides is shown in Fig. 5.5 (peptides were detected by UV absorbance at 214 nm). Again, the sharpest and cleanest peaks were selected for sequencing experiments and these are marked as fractions 12 and 20 on the profile.

A second clostripain digestion of the pure pyruvate kinase, performed for only 2 hr at 37°C, yielded the H.P.L.C. peptide profile shown in Fig. 5.6. In this profile, two major peaks (23 and 24) are clearly visible and these were collected for sequencing. It is possible, after such a short reaction

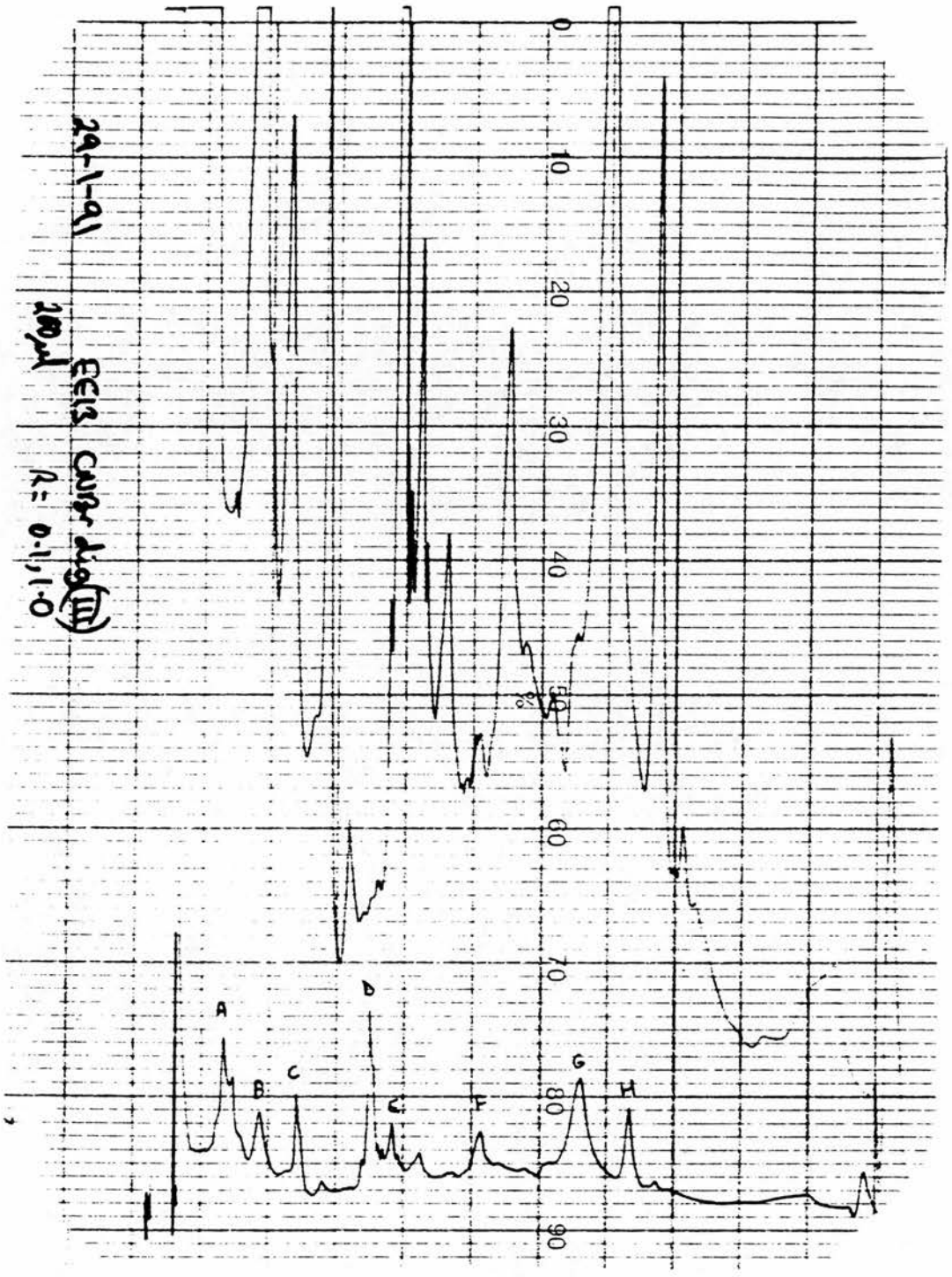


Figure 5.3: Separation of cyanogen-bromide digest-derived peptides using the Aquapore RP-300 column. Peaks labelled were collected in the smallest possible volumes from the column. Peptide 'D' was chosen for further characterisation. Peptide elution was monitored at 210 nm (full scale deflection on the graph shown was 0-0.1 AU).

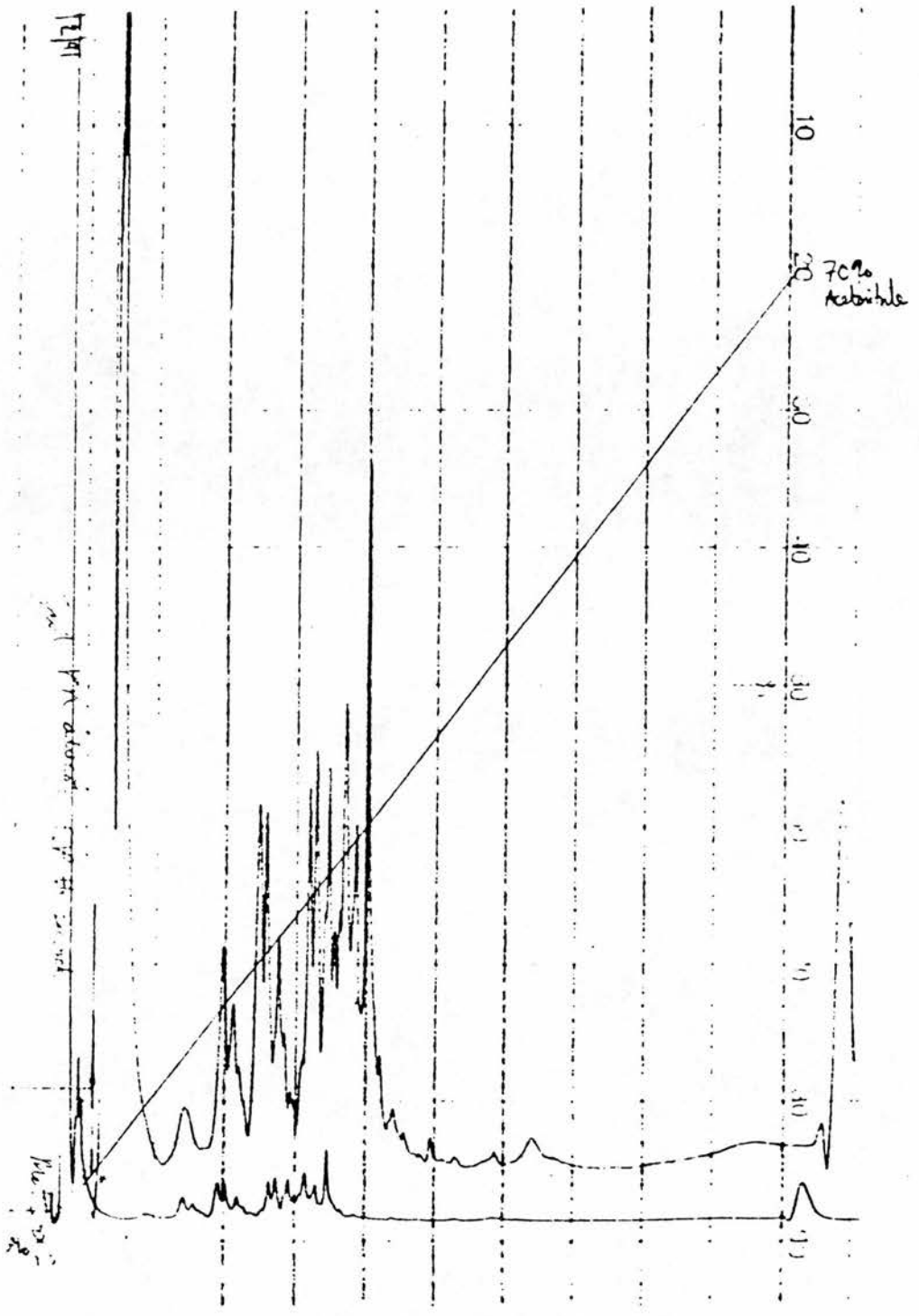


Figure 5.4: Separation of clostripain digestion-derived peptides using the Aquapore RP-300 column. The peaks were considered to be too poorly resolved for collection from this column. Peptide elution was monitored at 210 nm (full scale deflection on the graph shown was 0-0.1 AU).

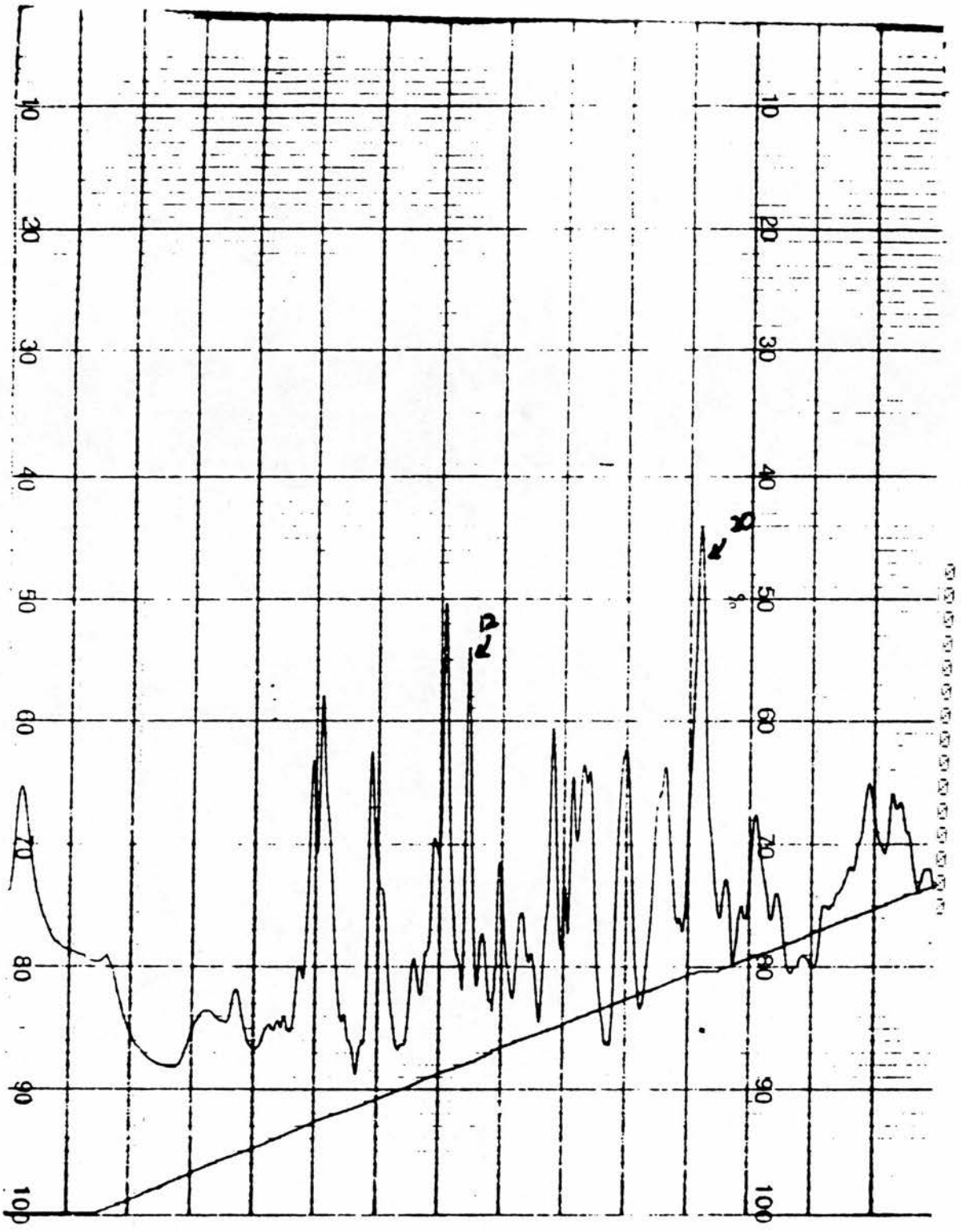


Figure 5.5: Separation of clostripain digestion-derived peptides using the PEP-RPC FPLC column. Fractions of 1 ml were automatically collected. The peptides labelled 12 and 20 were selected for further characterisation. Peptide elution was monitored at 214 nm (full scale deflection on the graph shown was 0-0.2 AU).

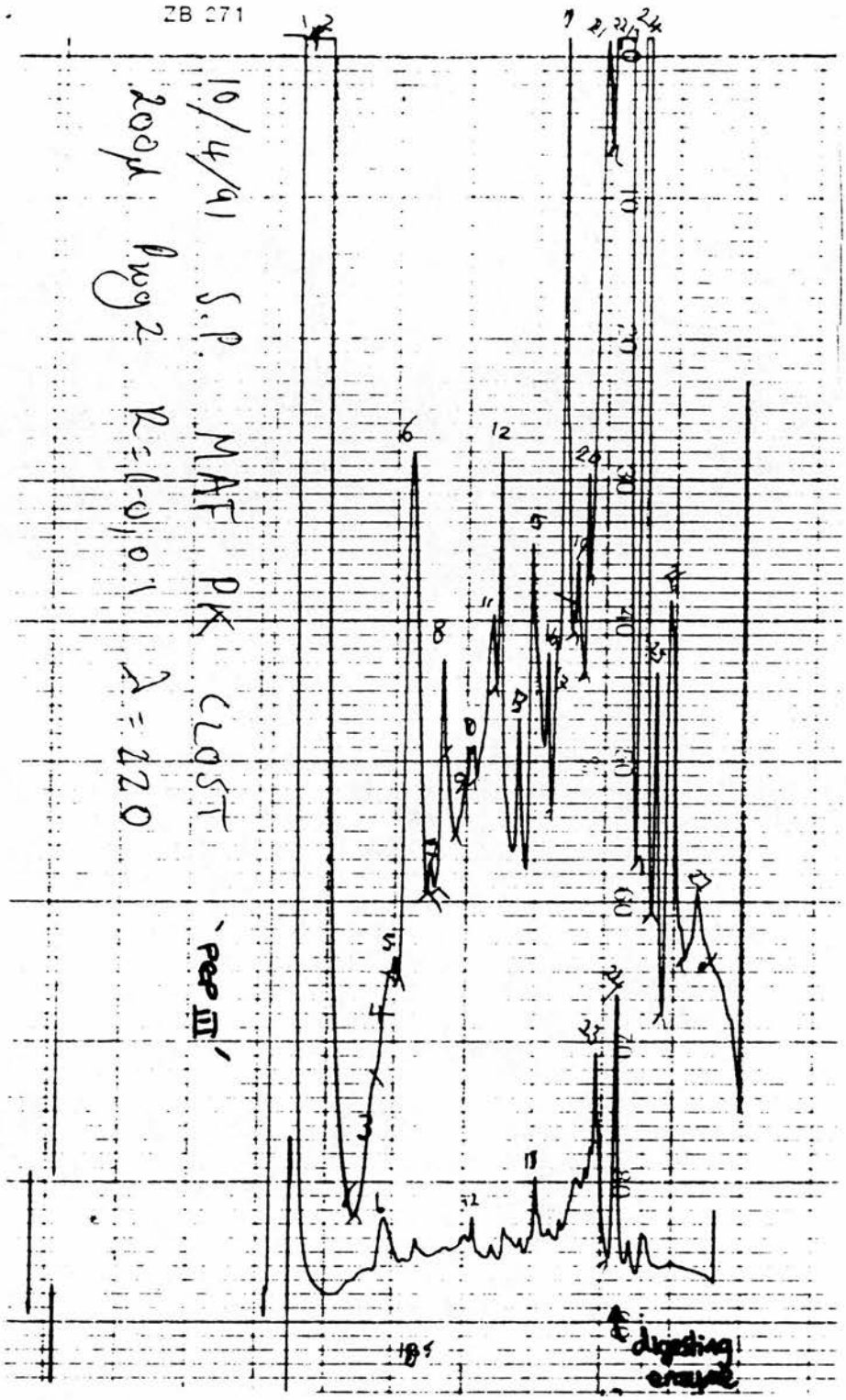


Figure 5.6: Separation of clostripain digestion-derived peptides (second, shorter digest) using the Aquapore RP-300 column. Peptide elution was monitored at 210 nm (full scale deflection in the graph shown was 0-0.1 AU).

time, that the digestion had not gone to completion in this case, a fact which may explain the different profile obtained for this reaction.

5.2.4 Sequencing of the peptides

Prior to its sequencing, each of the peptides selected above was subjected to amino acid analysis. This was simply to ensure that the peak collected did indeed contain proteinaceous material and was not a column artifact or a non-protein impurity. The qualitative amino acid analysis results for each of the peptides are collected together in Table 5.4. In every case, the peaks proved to contain peptide fragments and hence all were then submitted to the WELMET facility for microsequencing.

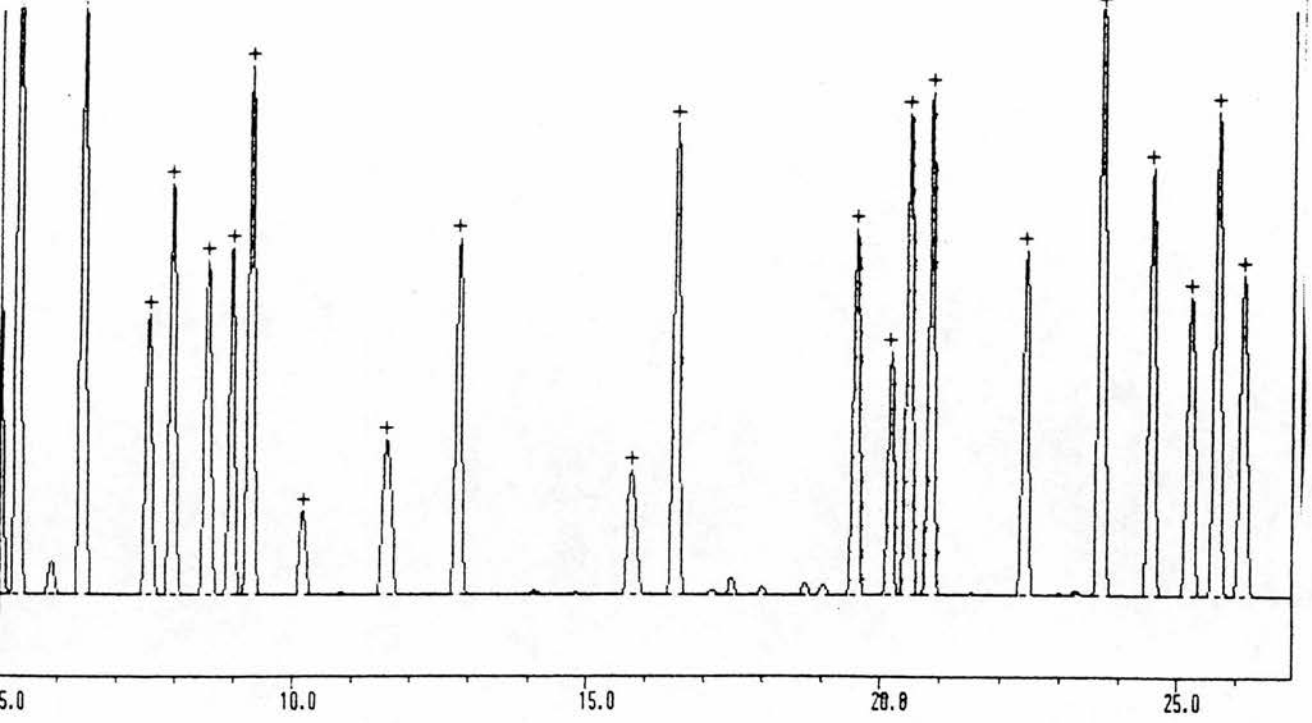
All of the sequencing described was done automatically on an Applied Biosystems 477A instrument. Loading of the sample was performed as described in 2.2.14 and, for each peptide, 300 pmol was used. The machine incorporates the chemistry of the PITC degradation procedure developed originally by Edman (Edman and Begg, 1967). The process is a cyclical one in which amino acid residues are cleaved one by one from the N-terminus of the peptide. There are three steps in each complete cycle: coupling of PITC with the amino terminal residue, cleavage of the amino terminal residue derivative via cyclization in acidic media, and conversion of the thiazolinone (Atz) derivative so formed to the more stable thiohydantoin (Pth) derivative. The 477A Derivatizer is linked to an on-line 120A phenylthiohydantoin analyser which collects the output of each cycle of the 477A instrument. It then separates out the derivatives chromatographically using a reverse phase H.P.L.C. column which is eluted with a linear gradient of acetonitrile. The derivatives are detected spectrophotometrically in the eluent of the column at 269 nm. Fig. 5.7 shows a calibration run of the sequencer with known concentrations of the Pth derivatives of each amino acid. The sequences of each of the peptides were then determined using this calibration as an internal standard.

5.2.4.1 Cyanogen bromide peptide p'D'

The sequence report for peak 'D' isolated from the cyanogen bromide digest of the pyruvate kinase was blank. Possible reasons for this result are discussed later.

amino acid	p12	p20	p23
B	✔	✔	✔
Z	✔	✔	×
S	✔	✔	✔
G	✔	✔	✔
H	✔	×	×
R	✔	×	×
T	✔	✔	✔
A	✔	✔	✔
P	✔	✔	✔
Y	✔	✔	✔
V	✔	✔	✔
M	✔	✔	×
C	×	×	×
I	✔	✔	✔
L	×	✔	✔
F	✔	✔	×
K	✔	✔	×

Table 5.4: Amino acid analysis results for peptides p12, p20 and p23. The purpose of the analysis (see text) was simply to confirm the proteinaceous nature of the peptides and to give some indication of their probable sequences and hence the table only shows which amino acids were present in each peptide. Precise pmol values were not available for the peptides due to a calibration error in the system. ✔ indicates presence of an amino acid residue. × indicates absence of amino acid residue. (✔ indicates amino acid subsequently assigned to peptide sequence).



PEAK TABULATION : (100% injection) Calibration : SDP291

Peak ID	R.Time (min)	C.Time (min)	Height (uAU)	Pmol	Peak ID	R.Time (min)	C.Time (min)	Height (uAU)	Pmol
	5.02		14127		CYS	20.15	20.15	10827	75.00
ASP	5.33	5.33	36345	75.00	MET	20.45	20.45	21555	75.00
	5.90		1494		VAL	20.82	20.82	22545	75.00
ASN	6.43	6.43	26289	75.00	DPT	22.42	22.42	15465	75.00
SER	7.55	7.55	12564	75.00	TRP	23.70	23.70	28170	75.00
GLN	7.95	7.95	18357	75.00	PHE	24.55	24.55	19194	75.00
THR	8.57	8.57	14958	75.00	ILE	25.22	25.22	13479	75.00
GLY	8.97	8.97	15453	75.00	LYS	25.67	25.67	21690	75.00
GLU	9.27	9.27	23631	75.00	LEU	26.12	26.12	14463	75.00
OMP	10.17	10.17	3768	75.00					
HIS	11.62	11.62	6870	75.00					
ALA	12.83	12.83	15912	75.00					
ARG	15.77	15.77	5577	75.00					
TYR	16.53	16.53	21123	75.00					
PRO	19.55	19.55	16377	75.00					

Figure 5.7: Calibration of the Applied Biosystems 477A protein sequencer. A standard mix of amino acids (75 pmol of each) was used to calibrate the column used in the machine for subsequent identification of eluted amino acid derivatives. Peak heights, corresponding pmol amounts and retention times are indicated for each amino acid derivative.

5.2.4.2 Clostripain peptide p12

The sequence report for p12 is shown in Table 5.5. The deduced sequence is **KAMDVAEVR**. The initial yield for the peptide was 195.8 pmol and the repetitive yield for alanines 2 and 6 was 72% (for valines 5 and 8 it is 76%). 8 amino acids which were detected in the amino acid analysis of the peptide are not found in this sequence. This strongly suggests the presence of contaminating peptide(s). The sequence(s) of this (these) peptide(s) may be interpretable from a closer analysis of Table 5.5 although the mixture appears to be quite complex.

5.2.4.3 Clostripain peptide p20

The sequence report for p20 is shown in Table 5.6. The deduced sequence is unclear up to residue 4 but is clearly **AVALDTK** from there. After the lysine residue the yield of each step is markedly reduced indicating that clostripain has probably cut after this residue. It seems likely, therefore, that the succinylation reaction did not go to completion and that this particular lysine residue was not protected from proteolytic action. The initial yield for the peptide was 208.95 pmol and the repetitive yield for alanines 4 and 6 was 82%. Again the presence of contaminating peptide(s) is suggested by the fact that 6 amino acids detected by the analysis are not present in the peptide sequence. Most contaminant appears in the first two cycles - this may be a carry-over from a previous sequencing experiment.

5.2.4.4 Clostripain peptide p23

The sequence report for p23 is shown in Table 5.7. The deduced sequence is **AGIYLPGANT**. The initial yield of the peptide was 193.63 pmol and the repetitive yield for glycines 2 and 7 was 61%. Since the peptide sequence does not end in either an arginine or a lysine residue (the cut sites of clostripain) it is possible that this peptide may represent the carboxyl terminus of the pyruvate kinase. This is discussed further, together with the other sequencing results in the next section. Only 3 amino acids are present in the analysis which do not appear in the peptide most of which appear in only the first cycle. This was the easiest of the sequences to interpret.

	Cycle										
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
Ala	86	**522**	313	0	5	**137**	0	6	4	1	5
Arg	3	0	0	0	0	1	14	18	35	0	0
Asn	10	0	50	25	0	0	2	0	0	0	0
Asp	0	0	14	**183**	0	1	6	0	0	0	0
Cys	1	1	1	1	1	1	1	1	1	1	1
Glu	1	21	0	1	14	27	**178**	0	0	0	1
Gln	0	0	0	1	0	2	0	0	0	0	0
Gly	0	17	0	5	13	42	10	3	1	2	5
His	1	0	1	7	17	48	0	1	0	0	1
Ile	51	0	0	4	0	1	0	1	0	0	0
Leu	4	0	2	0	2	1	0	3	0	0	0
Lys	**196**	0	0	0	0	5	2	1	0	0	0
Met	0	0	**132**	0	0	0	0	0	0	0	0
Phe	4	1	0	0	12	3	0	0	0	0	0
Pro	0	3	21	38	0	0	0	0	0	0	0
Ser	38	0	0	1	1	2	0	0	0	0	0
Thr	10	0	1	1	0	4	6	0	0	0	0
Trp	10	6	0	0	0	0	1	1	1	1	1
Tyr	3	0	1	1	1	12	0	1	1	1	1
Val	13	22	17	192	**529**	0	19	**233**	0	5	5

Table 5.5: Sequence report for peptide p12. The values shown are lag corrected in pmol and have been rounded up or down to the nearest integer. Some cycles suggest that contaminating peptides may be present (see text).

	Cycle										
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
Ala	56	0	12	**178**	34	**120**	11	14	0	2	1
Arg	0	0	1	0	1	0	0	0	0	0	0
Asn	0	209	1	7	0	6	0	0	0	0	0
Asp	2	7	0	2	1	0	2	**46**	10	4	2
Cys	0	0	0	0	0	0	0	0	0	0	0
Glu	0	1	14	22	2	4	1	1	1	1	1
Gln	0	1	3	0	1	0	0	0	0	0	5
Gly	20	41	3	0	0	27	3	0	0	0	0
His	0	1	0	0	0	0	0	0	0	1	1
Ile	51	24	1	0	2	0	0	0	1	1	0
Leu	0	30	1	1	3	1	**105**	14	10	0	0
Lys	2	0	0	0	0	0	2	0	0	**9**	0
Met	16	0	1	4	0	0	0	0	1	0	0
Phe	1	3	25	4	0	0	0	0	7	0	1
Pro	0	209	32	5	0	21	1	0	0	0	0
Ser	31	4	3	11	1	0	10	0	1	0	0
Thr	8	6	0	0	11	0	9	1	**39**	5	0
Trp	3	9	0	1	1	0	0	0	0	5	2
Tyr	3	0	0	0	1	10	0	0	0	1	1
Val	57	1	**116**	21	**147**	16	0	3	0	0	2

Table 5.6: Sequence report for peptide p20. The values shown are lag corrected in pmol and have been rounded up or down to the nearest integer. The first two cycles are not interpretable and may be contaminated.

	Cycle										
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
Ala	21	**25**	0	4	0	0	0	1	**13**	0	3
Arg	0	0	0	0	1	0	0	0	0	0	0
Asn	3	0	0	0	3	0	3	1	0	**9**	0
Asp	0	2	0	0	0	5	0	4	1	1	0
Cys	1	0	0	0	1	0	0	0	0	0	0
Glu	0	0	5	0	0	0	0	0	0	2	0
Gln	0	0	0	0	0	2	0	0	0	0	0
Gly	0	0	**19**	5	0	2	0	**15**	1	0	1
His	1	0	0	0	0	1	0	0	0	0	0
Ile	33	3	0	**16**	0	4	2	0	0	0	0
Leu	11	3	0	0	2	**18**	0	0	0	4	1
Lys	4	0	1	0	0	0	0	0	0	0	0
Met	1	0	0	0	0	0	0	0	0	0	0
Phe	3	0	0	0	1	0	0	0	0	0	0
Pro	6	0	1	1	3	0	**9**	0	0	0	1
Ser	45	0	2	0	3	2	0	0	4	0	0
Thr	10	0	0	3	0	5	0	5	0	1	**7**
Trp	0	6	5	3	0	0	0	5	0	1	2
Tyr	2	0	0	0	**10**	2	0	1	0	0	0
Val	10	6	0	0	5	1	0	5	1	1	0

Table 5.7: Sequence report for peptide p23. The values shown are lag corrected in pmol and have been rounded up or down to the nearest integer. The first cycle is not interpretable and may be contaminated.

5.3 Discussion

5.3.1 Amino acid analysis

The amino acid composition of the *T. acidophilum* pyruvate kinase is shown in Table 5.3 together with the compositions of the counterpart enzymes isolated from chicken muscle and *Bacillus stearothermophilus* for comparison. Some comparative features interpreted from this table are worthy of note. The levels of glycine, glutamate and proline residues seem to be elevated in the archaebacterial enzyme compared to those found in the other enzymes. The level of lysine residues seems, in contrast, to be reduced in the archaebacterial enzyme. These particular observations are of interest as these features seem to be generally applicable to thermostable proteins (Watson and Littlechild, 1990).

The high levels of glycine and proline residues may be explained by the ability of these amino acids to help form tighter (glycine) or less flexible (proline) main chain conformations of the protein. These residues may, therefore, contribute towards protein stability at high temperatures by reducing the entropy of the unfolded state and thereby making its formation less desirable energetically. The correlation between enhanced stability at elevated temperatures and high levels of proline residues in the protein structure has recently been formalised as a hypothesis by Professor Yuzuru Suzuki of Kyoto University. A study of oligo-1,6-glucosidases (dextrin-6-a-D-glucanohydrolase, EC 3.2.1.10) isolated from many sources led to the observation that the thermostability of a globular protein can be enhanced cumulatively by increasing the frequency of proline occurrence at the second sites of β-turns without significant alteration of either the backbone conformation or the protein's function (Suzuki, 1989).

The increased number of glutamate residues can be explained by the enhanced stabilisation of helices which can be a result of the replacement of uncharged residues by ones that have charge appropriate to their position in the helix. An example of this is the phosphoglycerate kinase (PGK) of *Bacillus stearothermophilus* in which 8 of the 13 helices are stabilised with respect to the comparative yeast enzyme by amino acid substitutions of this type (Davies *et al.*, 1991). The reduced number of lysine residues, again a general feature of thermophilic enzymes, is

probably due to their propensity to participate in degradative reactions at high temperatures. The residue level changes described here are over and above those seen for the *B. stearothermophilus* enzyme (another thermophilic pyruvate kinase) in comparison with the mesophilic chicken enzyme. This may be a reflection of the enhanced thermostability of the archaebacterial enzyme over the bacterial one.

The differences in amino acid compositions between the three pyruvate kinases compared here can be analysed in a more rigorous fashion by employing the Marchalonis and Weltman index (1971) according to the method of Cornish-Bowden (1977). The method provides a means of assessing the protein sequence identity between two proteins of approximately equal length (within 10%) using amino acid composition data. The index proposed ($S\Delta n$) is defined as half the sum of squares of the differences between the numbers of residues of each type in the two proteins. That is to say:

$$S\Delta n = 1/2 \times S\ (n_i A - n_i B)^2$$

where: $S\Delta n$ is the Marchalonis and Weltman index and $n_i A$ and $n_i B$ are the number of residues of the ith type of amino acids in proteins A and B respectively. For unrelated proteins of length N, the index is expected to exceed 0.42N in about 95% of comparisons. If the value of the index is less than 0.42N, then the proteins may show some sequence identity above the threshold random value (see chapter 1). In the case of the pyruvate kinases, the value of 0.42N is around 250.

$S\Delta n$ values were calculated for comparisons of the archaebacterial pyruvate kinase with both theB. stearothermophilus and chicken muscle enzymes. For the *Thermoplasma*/*Bacillus* comparison, the value of the index was 850. For the *Thermoplasma*/chicken muscle comparison, the value of the index was 1441. In both cases it is apparent that the value for the $S\Delta n$ index far exceeds the value of 0.42N and hence, from this statistical analysis, it seems unlikely that the archaebacterial enzyme will bear any sequence resemblance to its counterparts from eukaryotic and bacterial sources. A caveat is that this analysis has been employed by many groups in the past with varying degrees of success (although most errors have been in the prediction of sequence similarity between proteins which have later been shown by sequencing studies to be

completely unrelated in primary structure).

5.3.2 Amino acid sequences of peptides

The prediction resulting from the Cornish-Bowden analysis confirmed the importance of the next step in the characterisation of the *T. acidophilum* pyruvate kinase - the partial sequencing of the purified enzyme. As the amino-terminus of the enzyme proved to be chemically blocked (a property it has in common with many other pyruvate kinases, see Table 4.2), sequence was not obtained by this method. Instead, the pure archaebacterial enzyme was cleaved, by both chemical and enzymatic methods, into a series of small peptides which were then chromatographically separated and subjected to Edman degradation sequencing. Four peptides were chosen for sequencing, a cyanogen bromide peptide (p'D') and three clostripain derived peptides (p12, p20 and p23).

5.3.2.1 Cyanogen bromide peptide p'D'

No sequence was obtained from this peptide. There are two possible explanations for this observation. First, the peptide picked may in fact represent the amino-terminus of the full protein. Secondly, the amino-terminal residue of the peptide may have been a glutamine. If this was the case, a side reaction could have occurred during the cleavage process resulting in the cyclisation of the released amino-terminal glutamine to a pyrrolidonecarboxylyl residue which would, in effect, have blocked the end of the peptide and prevented its sequencing.

5.3.2.2 Clostripain peptides p12, p20 and p23

The deduced amino acid sequence for these peptides were **KAMDVAEVR, AVALDTK** and **AGIYLPGANT** respectively. Possible alignments for the peptides to other pyruvate kinase sequences are shown in Fig. 5.8. Whilst the match for p12 to the alignment is not conspicuously good, in the case of p20, the alignment has identified a region of the pyruvate kinase sequence which is completely conserved evolutionarily. This result would seem to suggest first, that this region is important for the activity of the enzyme and secondly, that the archaebacterial enzyme retains some sequence homology with its eukaryotic and bacterial counterparts, at least in functionally important

regions. This is despite the conclusion of the Cornish-Bowden analysis that the enzyme as a whole should present little or no sequence identity to other pyruvate kinases. The evidence of the p20 sequence may be supported by the sequence of p23 which shows some similarity to the carboxy-terminal region of the trypanosomal pyruvate kinases. The possibility that p23 is C-terminal in nature is also suggested by the fact that its sequence does not end in a lysine or arginine residue. Two alignments are presented for p23 as the alignment program GAP produced two possible regions of similarity in the pyruvate kinase sequence alignment to this peptide.

5.4 Summary of conclusions

The amino acid analysis of the *T. acidophilum* pyruvate kinase suggests that the protein may employ well established methods of protein stabilisation at high temperatures. Its differing levels of some amino acid residues are typical of other proteins adapted to thermophilic environments. A Cornish-Bowden analysis of the composition implies that little or no sequence identity will be found between the sequence of the archaebacterial enzyme and those isolated from eukaryotic and bacterial sources.

The partial amino acid sequences obtained, however, may suggest that the protein does maintain some sequence homology with non-archaebacterial pyruvate kinases, especially in regions of functional importance such as the active site. These predictions were useful in determining the strategy employed for the cloning and sequencing of the *T. acidophilum* pyruvate kinase gene which is fully described in the following chapter.

```
Peptide p20      A V A L D T K
                                                          124
hummus2        V A V A L D T K G P E I R T G L I K
ratmus2        V A V A L D T K G P E I R T G L I K
ratmus1        V A V A L D T K G P E I R T G L I K
catmus1        V A V A L D T K G P E I R T G L I K
chimus         V A I A L D T K G P E I R T G L I K
humliv         V A I A L D T K G P E I R T G I L Q
ratliv         V A I A L D T K G P E I R T G V L Q
ratrbc         V A I A L D T K G P E I R T G V L Q
potCy          C A V M L D T K G P E I R T G F L T
Anid           V A I A L D T K G P E I R T G N T V
Anig           L A I A L D T K G P E I R T G N T P
yea            L A I A L D T K G P E I R T G T T T
TbrCy1         I G I A L D T K G P E I R T G L F K
TbrCy2         I G I A L D T K G P E I R T G L F K
Eco            A A I L L D T K G P E I R T M K L E
Bst            V A I L L D T K G P E I R T H N M E
                       * * * * * * * * * *
```

Figure 5.8(i): **Alignment of peptide p20 to pyruvate kinase sequences.**

```
Peptide p12      K A M D V A E V R

                                                                  260
hummus2          R K A S D V H E V R K V L G E K
ratmus2          R K A A D V H E V R K V L G E K
ratmus1          R K A A D V H E V R K V L G E K
catmus1          R K A S D V H E V R K V L G E K
chimus           R K A A D V H A V R K V L G E K
humliv           R K A S D V A A V R A A L G P E
ratliv           R K A S D V L A V R D A L G P E
ratrbc           R K A S D V L A V R D A L G P E
potCy            R K G S D L V N V R K V L G P H
Anid             R R G S D I R H I R E V L G E E
Anig             R R G S D I R H I R E V L G E E
yea              R T A N D V L T I R E V L G E Q
TbrCy1           R T A E Q V R E V R A A L G E K
TbrCy2           R T A E Q V R E V R A A L G E K
Eco              R K R S D V I E I R E H L K A H
Bst              R R A S D V L E I R E L L E A H
                 *                 *
```

Figure 5.8(ii): Alignment of peptide p12 to pyruvate kinase sequences.

```
peptide p23          A G I Y L P G A N T

hummus2          T M R V V P V P
ratmus2          T M R V V P V P
ratmus1          T M R V V P V P
catmus1          T M R V V P V P
chimus           T M R V V P V P
humliv           I M R V L S I S
ratliv           I M R V L S V S
ratrbc           I M R V L S V S
potCy            K I C V V K
Anid             T V R V V P A E E N L G
Anig             T V R V V P A E E N L G
yea              T L Q V S T V
TbrCy1           Q T R L I Y L P
TbrCy2           Q T R L I Y L P
Bst              N L M K V H V I S D L L

peptide p23        A G I Y L P G A N T
                                      216
hummus2          K K G V N L P G A A V D
ratmus2          K K G V N L P G A A V D
ratmus1          K K G V N L P G A A V D
catmus1          K K G V N L P G A A V D
chimus           K K G V N L P G A A V D
humliv           R K G V N L P G A Q V D
ratliv           R K G V N L P N T E V D
ratrbc           R K G V N L P N T E V D
potCy            R K N V N L P G V V V D
Anid             R K G V N L P G T D V D
Anig             R K G V N L P G T D V D
yea              H K G V N L P G T D V D
TbrCy1           R R G I N L P G C E V D
TbrCy2           R R G I N L P G C E V D
Eco              N K G V N L P G V S I A
Bst              K K G V N V P G V K V N
                         *   *
```

Figure 5.8(iii): Alignment of peptide p23 to pyruvate kinase sequences.

Chapter 6: Cloning and Sequencing of *T. acidophilum* Genomic DNA

CHAPTER 6: CLONING AND SEQUENCING OF *THERMOPLASMA ACIDOPHILUM* DNA

6.1. Introduction

6.1.1. Design of probes for the *T. acidophilum* pyruvate kinase gene.

The peptides derived from clostripain digestion of the purified *T. acidophilum* pyruvate kinase (chapter 5) were used as a basis for the designing of oligonucleotide probes for the pyruvate kinase gene (Fig. 6.1). There are two criteria for designing gene probes from amino acid sequences: maximum length and minimum complexity (complexity being due to 3rd base 'wobbles' introduced by the degeneracy of the genetic code). Minimum complexity is an advantage as this will reduce the chance of non-specific binding of the probe to other sequences of the DNA. This goal was achieved by employing a 'best guess' approach using the codon preference table for *T. acidophilum* (Table 6.1) which had been derived from the study on the citrate synthase gene of that organism (Sutherland, 1991).

Three probes were designed and used for the purpose of isolating and sequencing the pyruvate kinase gene (see Fig. 6.1). Peptide p12 was not used for probe design experiments as, of the sequences obtained, that of p12 was the most difficult to interpret accurately (see Table 5.5). Probe 1 (O1) was based on the sequence of the peptide p20. Since this peptide was identified as corresponding to a completely conserved region of all sequenced pyruvate kinases, it was decided to assume that this region would also be present in its entirety in the *T. acidophilum* pyruvate kinase sequence. Hence, O1 was designed complementary to this entire conserved region (amino acids 107-120, numbered according to the alignment shown in Fig. 1.7). The resultant oligonucleotide was a redundant 32-mer taking into account only those bases in the wobble positions which were suggested by the codon usage table. Probe 2 (O2) was based on the sequence of peptide p23 - a redundant 27-mer again based on the codon usage table.

Probe 3 (O4) was based on another highly conserved region of pyruvate kinases as determined from the sequence alignment (Fig. 1.7).

F	TTT	3	**S**	TCT	3	**Y**	TAT	6	**C**	TGT	0
F	TTC	12	**S**	TCC	5	**Y**	TAC	14	**C**	TGC	0
L	TTA	0	**S**	TCA	5	*	TAA	0	*	TGA	1
L	TTG	1	**S**	TCG	2	*	TAG	0	**W**	TGG	4
L	CTT	5	**P**	CCT	2	**H**	CAT	2	**R**	CGT	0
L	CTC	8	**P**	CCC	2	**H**	CAC	4	**R**	CGC	0
L	CTA	1	**P**	CCA	7	**Q**	CAA	0	**R**	CGA	1
L	CTG	9	**P**	CCG	6	**Q**	CAG	10	**R**	CGG	0
I	ATT	4	**T**	ACT	3	**N**	AAT	3	**S**	AGT	0
I	ATC	8	**T**	ACC	6	**N**	AAC	12	**S**	AGC	4
I	ATA	19	**T**	ACA	5	**K**	AAA	4	**R**	AGA	7
M	ATG	12	**T**	ACG	7	**K**	AAG	26	**R**	AGG	11
V	GTT	4	**A**	GCT	7	**D**	GAT	10	**G**	GGT	5
V	GTC	6	**A**	GCC	10	**D**	GAC	9	**G**	GGC	10
V	GTA	5	**A**	GCA	19	**E**	GAA	12	**G**	GGA	6
V	GTG	7	**A**	GCG	8	**E**	GAG	18	**G**	GGG	5

Table 6.1: Codon usage table for the *T. acidophilum* citrate synthase gene.

OLIGO O1: $^{5'}$GCACTCGATACGAAGGGCCCAGAGATAAGGAC$^{3'}$

C G A G A

T

AA SEQUENCE: **AVALDTKGPEIRT**

OLIGO O2: $^{5'}$GCAGGCATATACCTCCCAGGCGCAAAC$^{3'}$

C G C

AA SEQUENCE: **AGIYLPGAN**

OLIGO O4: $^{5'}$TACCAICGTTCCCCGCTAGAICCGCAICTCTATGG$^{3'}$

AA SEQUENCE: **MVARGDLGVEIP**

Figure 6.1: Nucleotide sequences of oligonucleotide probes O1, O2 and O4 and the corresponding amino acid sequences.

This approach was based on the assumption that the archaebacterial enzyme, having already been shown to retain some regions of sequence homology with other pyruvate kinases, would retain more than one such region. The region chosen for the third probe was that between amino acids 290-301 (Fig. 1.7). A non-redundant 35-mer with inosine at unassignable wobble positions was designed complementary to this stretch of sequence.In each case, calculated guesses were made at each of the wobble positions in order to reduce the complexity of the probes.

[Note: O3 was a probe based on the N-terminal sequence of the protein of unknown function (Ta20Kp) which was purified from *T. acidophilum*. These experiments are more fully described in Appendix 1].

6.1.2. Cloning strategies

In a ligation reaction between vector and insert DNA, it is generally the case that the vector religates itself at a higher frequency than it ligates to the insert DNA. This problem was overcome using two alternative strategies for the cloning into plasmids of the fragments of *T. acidophilum* DNA which were identified using the probes described above.

First, directional cloning was employed - this approach uses a vector that is cut with two different restriction endonucleases such that it cannot religate to itself. A 2kb EcoRI/HinDIII *T. acidophilum* genomic DNA fragment was cloned in this way. Secondly, two of the probes described above (O1 and O4) were used as primers to generate a 0.6kb PCR product which was cloned using the TA cloning system [Invitrogen Corporation]; this system employs a vector (pCRII) which has 3' -end T overhangs such that it should not religate to itself. These probes were chosen as it was possible to predict, from the pyruvate kinase multiple sequence alignment (Fig. 1.7), the size and orientation of the product of the PCR reaction. This would obviously facilitate the alignment of the amino acid sequence derived from the PCR product to the pyruvate kinase sequences stored in the database. Since the position of the sequence corresponding to O2 (peptide p23) could not be definitively determined in the sequence, this probe was not used as a primer for PCR reactions.

6.2 Results

6.2.1 Preparation of *T. acidophilum* genomic DNA

DNA was prepared using a method based on that of Bowen *et al.* (1988) for *Thermus thermophilus* (see 2.2.19). A two litre batch of *T. acidophilum* culture yielded approximately 0.7mg of genomic DNA. When analysed on an agarose gel (Fig. 6.2), the DNA ran as a tight band at a size equivalent to uncut Lambda DNA, indicating it to be of high molecular weight. When digested with a restriction endonuclease such as EcoRI (Fig. 6.2), the DNA was cut to give distinctive and repetitive banding patterns, indicating that it was pure.

6.2.2 Preparation of Southern blots of restricted *T. acidophilum* DNA

A number of single and/or double digests of the *T. acidophilum* genomic DNA were carried out with the hexanucleotide- specific restriction endonucleases EcoRI, BamHI and HinDIII. The restriction digests were performed as described in section 2.2.20. A single digest (EcoRI) is shown in Fig. 6.2. Two double digests (EcoRI/HinDIII and EcoRI/BamHI) were run on the same gel and are shown in Fig. 6.3. The two gels were blotted separately on to Hybond N filters (see section 2.2.22) and the blots were hybridised with the O1 and O2 oligonucleotide probes as described below.

6.2.3 Use of probes O1 and O2

The two Southern blots described in the above section were hybridised with fluorescein-labelled probe O1 at 42°C. The autoradiograph obtained for the blot of the single enzyme digest after development is shown in Fig. 6.4 and that obtained for the blot of double enzyme digests is shown in Fig. 6.5. The blots were then stripped of O1, re-hybridised with probe O2 and the blots developed again. In the case of both blots, O2 identified exactly the same fragments as did O1. In the blot shown in Fig. 6.4, both probes hybridised to a fragment of size 7 kb. In the blot shown in Fig. 6.5, both probes hybridised to a EcoRI/HinDIII fragment of 2 kb and an EcoRI/BamHI fragment of size 4.5 kb.

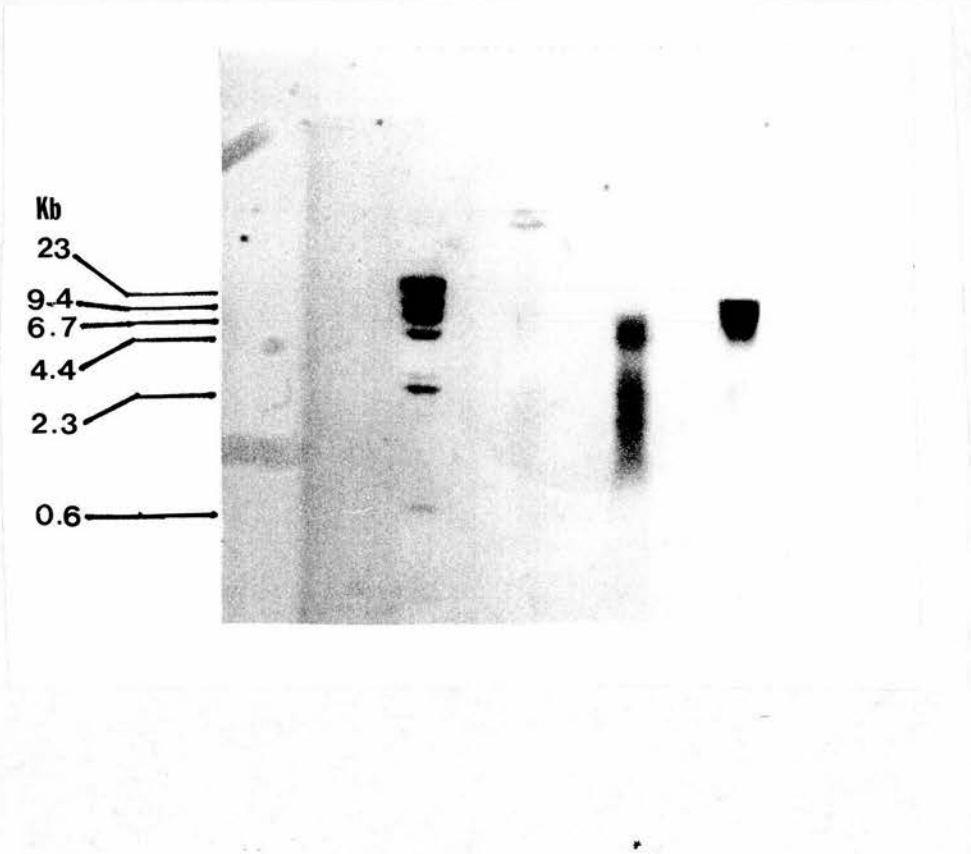


Figure 6.2: Analysis of *T. acidophilum* genomic DNA on a 1% (w/v) agarose gel. Lanes (L-R): Lambda DNA cut with HinD111; *T. acidophilum* genomic DNA restricted with EcoRI; *T. acidophilum* genomic DNA.

Figure 6.3: Analysis of *T. acidophilum* genomic DNA restricted with EcoR1/HinD111 and EcoR1/BamH1 enzymes. Lanes (L-R): Lambda DNA cut with HinD111; EcoR1/BamH1 double digest of *T. acidophilum* DNA; EcoR1/HinD111 double digest of *T. acidophilum* DNA.

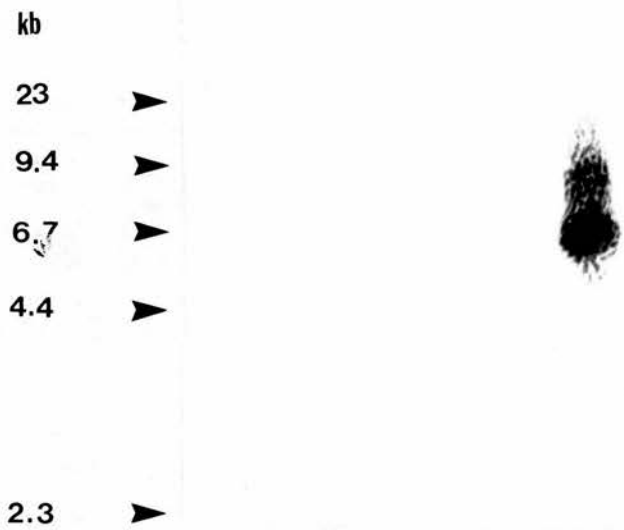


Figure 6.4: Hybridisation of single enzyme digest of *T. acidophilum* DNA with fluorescein-labelled probe O1. The gel used for the blot here was equivalent to that shown in Fig. 6.2 except that uncut genomic DNA was not included. Arrows in the left hand margin indicate the migration points of Lambda DNA-HinD111 fragments.

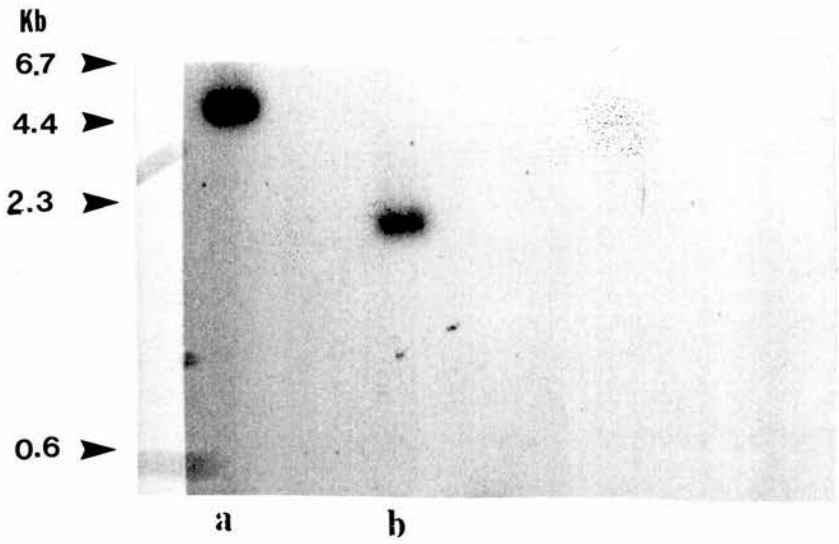


Figure 6.5: Hybridisation of double enzyme digests of *T. acidophilum* DNA with fluorescein-labelled probe O1. The gel used for this blot is described in Fig. 6.3. Arrows in the left hand margin indicate the migration points of Lambda DNA-HinD111 fragments. The tracks are labelled as in Fig. 6.3. The singly cut EcoRI fragment is 6.5kb but single digestion controls should have been included using BamHI and HinDIII separately.

6.2.4 Use of probes O1 and O4 as PCR primers

According to the sequence alignment of known pyruvate kinase primary structures (Fig.1.7), the region between the conserved sequence stretches which were used to design probes O1 and O4 is composed of 190 amino acids. If no introns were present in the coding sequence, this would correspond to a nucleic acid sequence of approximately 0.6kb. The probes O1 and O4 were used as primers for a PCR reaction (see 2.2.24) using *T. acidophilum* genomic DNA as the template and 45°C as the annealing temperature (the PCR was performed using an automated temperature cycling device programmed for 30 cycles of denaturing, annealing and extension reactions). After the reaction was complete, the products were analysed on an agarose gel (Fig. 6.6). Multiple (presumably mostly non-specific) products resulted although the major band was around 0.6kb in size. The PCR reaction was subsequently repeated using an annealing temperature of 55°C. Agarose gel analysis of the reaction products (Fig. 6.7) showed that only the approximately 0.6kb band remained indicating that the other products of the first reaction were indeed the result of some non-specific priming which was eliminated by the more stringent conditions of the second reaction.

6.2.5 Cloning and sequencing of the 2kb EcoRI/HinDIII fragment identified by the O1 and O2 oligonucleotide probes

The 2kb EcoRI/HinDIII fragment, identified by both O1 and O2, was selected for cloning because it was smaller than the EcoRI/BamHI fragment identified (and, therefore, easier to clone and quicker to sequence) whilst being large enough to possibly contain the entire gene sequence of the pyruvate kinase. This would be a particularly valid observation if the sequence of O2 (peptide p23) did prove to represent the carboxy-terminus of the pyruvate kinase (suggested in Fig. 5.8), as the sequences corresponding to O1 and O2 - which both hybridised to the same DNA fragment - would then represent both 'ends' of the pyruvate kinase coding region.

Approximately 25µg of *T. acidophilum* genomic DNA was digested first with EcoRI then with HinDIII and then separated on a 1.0% (w/v) low melting point (LMP) agarose gel. DNA of fragment size 2kb was

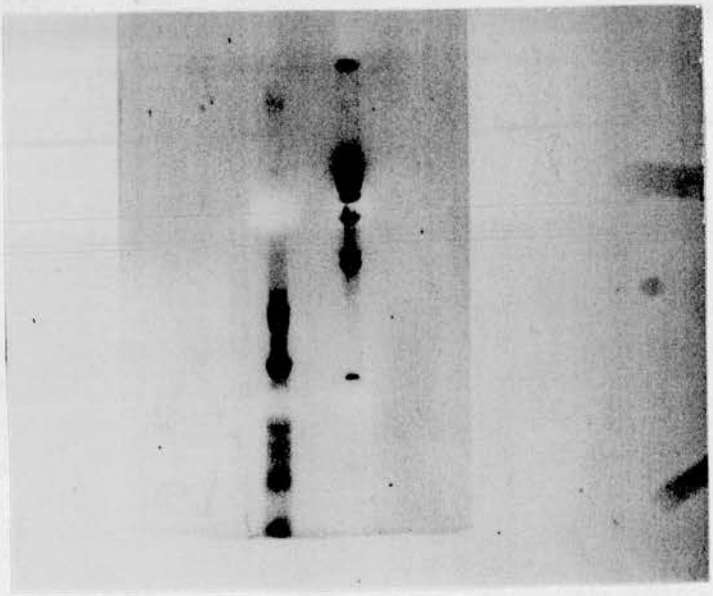


Figure 6.6: Analysis of products of the 45°C PCR reaction performed with probes O1 and O4 as primers. The products were separated using a 1% agarose gel. Lanes (1-2): Lambda DNA cut with HinD111; PCR products.

Figure 6.7: Analysis of product of the 55^{O}C PCR reaction performed with probes O1 and O4 as primers. The product was analysed on a 1% agarose gel. Lanes (L-R): Lambda DNA cut with HinD111; 0.6 kb PCR product.

recovered using the electrophoretic method of Pai and Bird (1992 - section 2.2.26) and purified using the 'Magic' DNA clean-up kit (Promega). The size selected DNA was then ligated into EcoR1/HinD111-cleaved pBluescript II vector (Stratagene - Fig. 6.8) and transformed into competent *E. coli* TG1 cells. Approximately 200 transformants were transferred on to Hybond N nitrocellulose and hybridised with fluorescein-labelled O1. The filter was washed twice in 6XSSC containing 1% (w/v) SDS, once at room temperature and once at 42°C, then developed using the ECL detection system (Amersham). Twenty five colonies gave strong positive signals; one of these was picked and plasmid was prepared from it using the 'Magic miniprep' kit (Promega). Fig. 6.9 shows a Southern blot of 5 µl (of 50 µl) of a miniprep of plasmid from a positive white colony selected from the transformant plate. The blot was developed as described in section 2.2.23.2 after hybridisation with fluorescein-labelled O1. The plasmid band is clearly detected indicating that the transformation has been successful.

The insert was first sequenced with an M13 universal forward primer supplied by Promega using the Promega 'fmol' thermal cycle sequencing kit. The methods used are described in section 2.2.25. Due to time restrictions, only partial sequence was obtained from this insert. The first 453 nucleotides of the sequence are presented in Fig. 6.10. No part of this region of the insert corresponds to the DNA sequence predicted for either of the probes O1 and O2. However, an open reading frame (the corresponding putative amino acid sequence is also given in Fig. 6.10) has been identified. The ORF appears to terminate at the stop codon underlined in Fig. 6.10 although it is possible that this is the result of difficulty in reading the DNA sequence at that point of the gel.

More sequence was obtained from the 2kb insert using O2 as the primer for the fmol kit. The nucleotide sequence obtained proved to be another open reading frame and is shown in Fig. 6.11. The corresponding amino acid sequence is also shown. Again no hybridisation sites for either O1 or O2 can be seen in the sequence.

6.2.6 Cloning and sequencing of the 0.6kb PCR product

Probes O1 and O4 were used (as previously described) as primers for a PCR reaction using 55°C as the annealing temperature. The single

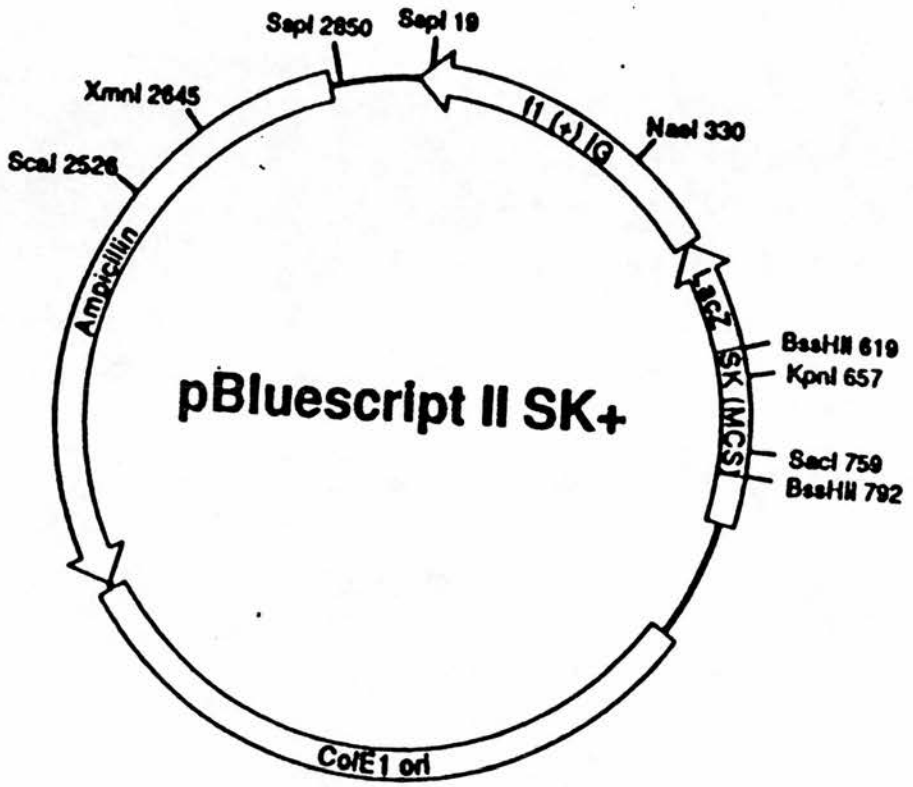


Figure 6.8: Map of plasmid pBluescript 11 SK+ (Stratagene). Multiple cloning site into which 2 kb EcoR1/HinD111 *T. acidophilum* genomic DNA fragment was inserted is indicated (MCS).

Figure 6.9: Hybridisation of plasmid preparation with labelled O1. The plasmid was prepared from a selected positive colony of TG1 cells after transformation with pBluescript 11 + 2 kb *T. acidophilum* DNA insert. Arrows in the left-hand margin indicate the migration points of Lambda DNA-HinD111 fragments. Lanes (L-R): Plasmid miniprep; total cell DNA preparation.

AAGCTTGCCCGGTGCAGTTATAGGTGGTATTTTTTACATTGCGCCCCT
GTACTATCTCCTCCGCCCTCTTGGTCTAACTCCTTCGGCGTGAGGTTA
CTTTATGATAGTCAAGACCGCCATAATAGCCACGACTACAAAAATAG
TAGGCGGTCAAGGTATTATATAGAAGACAATAAAAATAGCGTAGCGA
CTAGTAGGAACAAAAGAGGTTCGGCTTGCGAAGAAACTGTTACGCCA
CGGTACAAGGTCAAGGTCCCCGTAGCCCCTATACATACGGCCTTGGTA
TAAgcaacctaaacggaggcaagacctttttctccattgaaaggcagtagaagcgaa
gtataaggcagtgggcctttatccgtatccagcagacacctcccaagcttagagtatt
gctaccgtaagaagtgccaatagcagcgccattcgatagaccttgaac
ggcct 453

KLARCSYRWYFLHCAPVLSPPPSWSNSFGVRLLYDSQDRHNSHD
YKNSRRSRYYIEDNKNSVATSRNKRGSACEETVTPRYKVKVPVAP
IHTALV 95

Figure 6.10: Partial sequence of 2 kb EcoR1/HinD111 *T. acidophilum* genomic DNA insert in the pBluescript 11 vector. A stop codon is indicated (underlined). The polypeptide sequence derived from the open reading frame (285 nucleotides) is shown in bold below the nucleotide sequence. No obvious hybridisation site for either of the oligonucleotide probes (O1, O2) used can be found in this particular stretch of sequence.

AGTCGTGTGACCGCCGGCAATGATCACCAAGATCGAGCCATGGTTCGAACT

ACGTATGGAACTCATAAGATATCACAGTGGATTTATCGGAACCGCATTAG

TACCAGTATCGAC

Corresponds to sequence:

SRVTAGNDHQDRAMVRTTYGTHKISQWIYRNRISTSID

Figure 6.11: Partial sequence of 2kb p Bluescript II insert. The sequence was obtained using O2 as a primer for the fmol sequencing kit (Promega).

product of 0.6kb in size was recovered from an agarose gel by the electrophoretic method of Pai and Bird (1992, section 2.2.26) and was then ligated into the vector pCRII (Invitrogen - Fig. 6.12) and transformed into competent *E. coli* TG1 cells. Approximately 1000 transformants were transferred onto Hybond N and hybridised with fluorescein-labelled O1. The filters were washed in 6XSSC containing 1% (w/v) SDS once at room temperature and once at 42°C, and then developed using the ECL detection system (Amersham). Ten colonies gave strong positive signals. Plasmid was isolated from one of these colonies using a Magic miniprep column (Promega): a Southern blot of a miniprep of this purified plasmid probed with fluorescein labelled O1 is shown in Fig. 6.13. The blot was developed as described for the blot of the pBluescript II plasmid above. Again, the plasmid is clearly detected indicating that transformation has been successful. This insert was sequenced using both the Taq DyeDcoxy sequencing kit (Applied Biosystems) and the fmol kit. The methodology of the former system is also described in section 2.2.25.

Two Taq DyeDeoxy reactions were performed, one with M13 universal forward primer and the other with M13 universal reverse primer. Only the reaction with the forward primer was successful - the output from the ABI 373A instrument is shown in Fig. 6.14. The sequencing reactions using the reverse primer were unreadable due to excessive diffusion of the sample through the gel and time constraints prevented a repeat experiment. The sequence in Fig. 6.14 is clearly readable up to nucleotide 366 (after which unassignable bases are encountered with greater frequency) and is tabulated in Fig. 6.15 together with amino acid sequence corresponding to bases 116 - 366. Some amino acids in the sequence are unassignable and the resultant gaps have been marked with periods. [These gaps are due to the introduction of an uncalled base or an inappropriate stop codon into the nucleotide sequence. Given the probable position of the peptide within the pyruvate kinase sequence (see Chapter 7), it is highly unlikely that the peptide would terminate at this point. The phenomenon is probably due to inaccurate reading/copying of the template insert by *Taq* polymerase - an enzyme which is known to have an error rate of 0.25% for a 30 cycle amplification because it lacks an editing function. This is especially possible for this particular insert as it was also originally generated using the *Taq* enzyme. The most likely error that the enzyme would

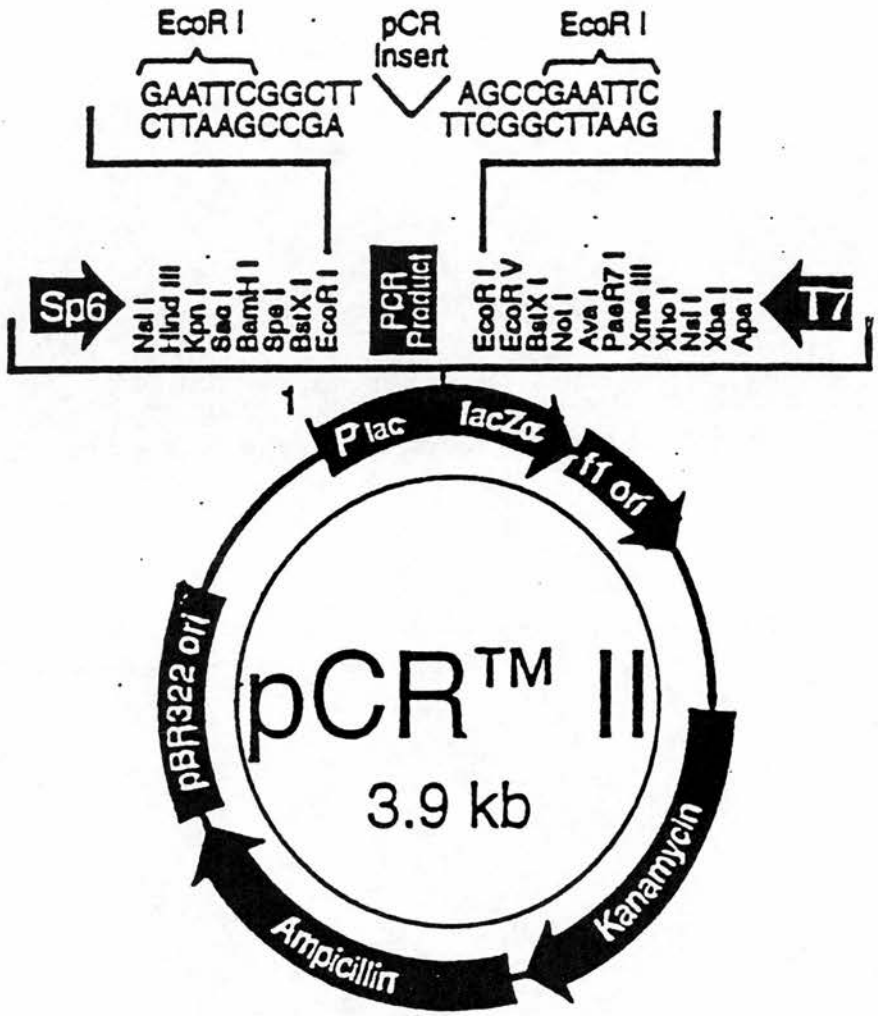


Figure 6.12: Map of plasmid pCR11 (Invitrogen). The sequence consists of: Lac Z gene (bases 1-571); Sp6 promoter (239-255), Multiple cloning site (269-381); T7 promoter (388-407); F1 origin (572-986); Kanamycin resistance (987-2114) and Ampicillin resistance (2133-2992).

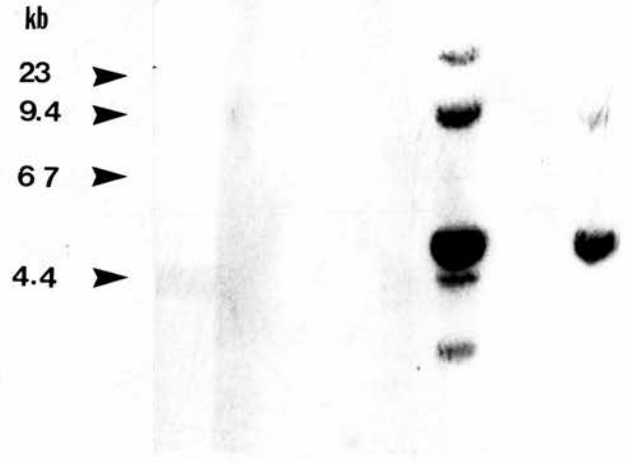


Figure 6.13: Hybridisation of plasmid preparation with labelled O1. The plasmid was prepared from a selected positive colony of TG1 cells after transformation with pCR11 + 0.6 kb PCR insert. Arrows in the left-hand margin indicate the migration points of Lambda DNA-HinD111 fragments. Lanes (L-R): Plasmid miniprep; total cell DNA preparation.

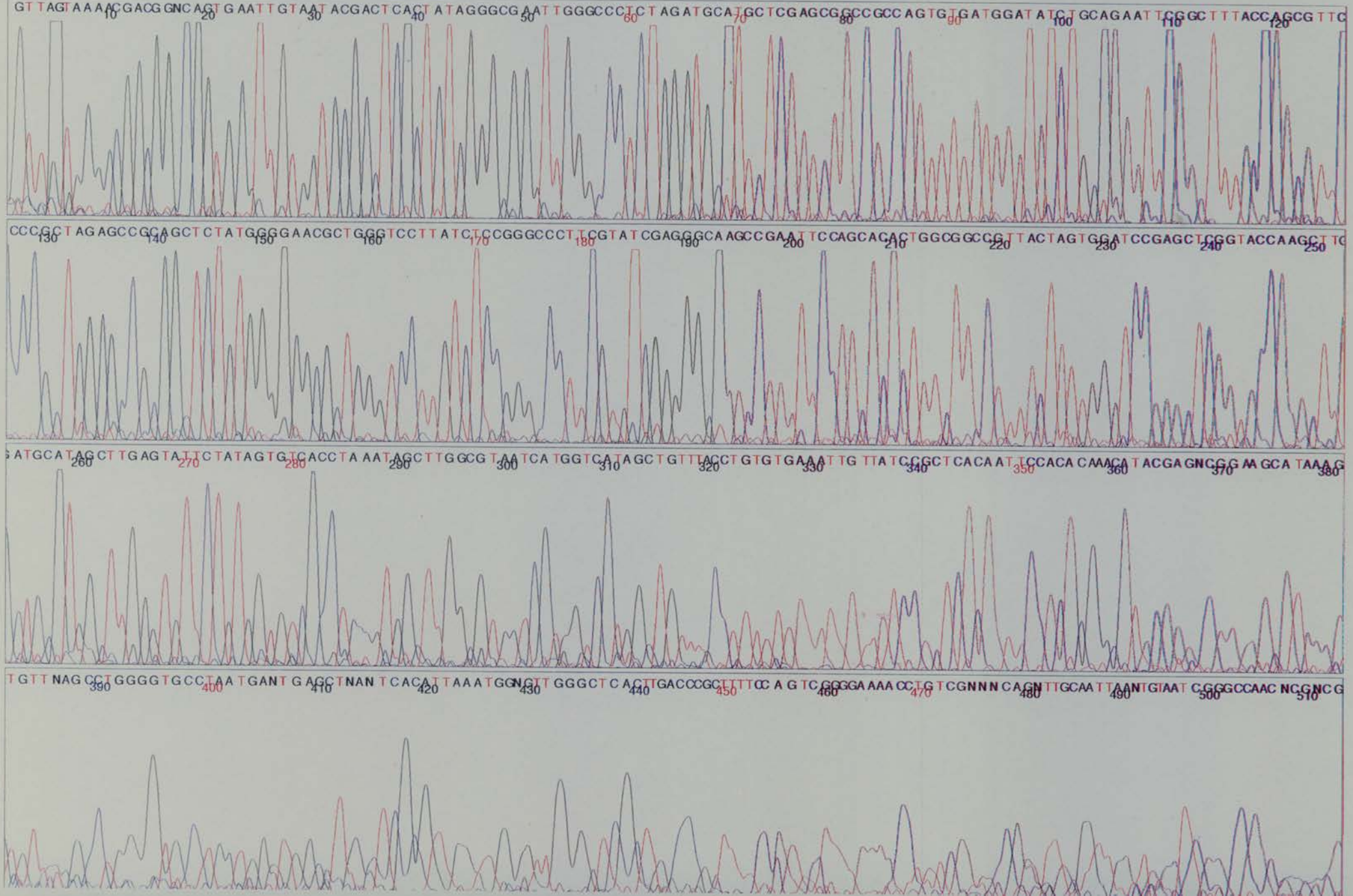
Figure 6.14: Output from the Applied Biosystems 373A DNA sequencer. Template DNA was pCRII vector transformed with the 0.6kb PCR product. Primer was M13 universal forward primer.

Model 373A
Version 1.1f

DyeTerminator{AnyPrimer}
Lane 5
Signal: G:55 A:24 T:23 C:18

9206653 MATRIX FILE
SIMON M13 F

X: 0 to 8659 Y: 0 to 1200



gttagtaaaacgacggncagtgaattgtaatacgactcactatagggcgaattgggccctct

agatgcatgctcgagcggccgccagtgtgatggatatctgcagaattcggcttTACCAGC

GTTCCCCGCTAGAGCCGCAGCTCTATGGGGAACGCTGGGTCCTTATCTCCGG

GCCCTTCGTATCGAGGGCAAGCCGAATTCCAGCACACTGGCGGCCGTTACT

AGTGGATCCGAGCTCGGTACCAAGCTTGATGCATAGCTTGAGTATTCTATA

GTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTACCTGTGT

GAAATTGTTATCCGCTCACAATTCCACACAAACATACGAGNCGGAAGCAT

AAAGTGTTNAGCCTGGGGTGCCTAATGANTGAGCTNANTCACATTAAATGG

NGTTGGGCTCACTTGACCCGCTTTTCCAGTCGGGGAAAACCTGTCGNNNCA

GNTTGCAATTAANTGTAATCGGGCCAACNCGNCG

Corresponds to sequence:

MVARGDLGVEIPLATQE.RPGKHSSRSA.GRVTAGNDHLGSSHGSNYVSNS.D

ITVDLSNRISTSIDKWTQFNNRRVLRCVCMLAEA.SVKPGVPN

Figure 6.15: Sequence of the 0.6kb PCR product inserted into pCRII obtained using M13 universal forward primer. Lower case nucleotides represent the sequence of the vector and upper case nucleotides represent the sequence of the insert. The sequence underlined corresponds to a region complementary to probe O4.

make is the substitution of one base for another and, hence, it has been assumed that the frame of reading has not been shifted by the errors. Instead, the entire codon has been omitted and a single amino acid gap introduced. These gaps have been taken into account in the attempts to align the peptide with other pyruvate kinase sequences (Chapter 7). Despite this difficulty, a clear hybridisation site for O4 can be seen between bases 116 - 151.

Only one fmol kit reaction was performed using the transformed pCRII vector as template with M13 universal reverse primer. The sequence obtained is shown in Fig. 6.16. A hybridisation site for O1 can be seen between bases 1 and 34.

6.3 Discussion

6.3.1 Use of probes O1 and O2

Two oligonucleotides, O1 and O2, based on the sequences of internal peptides obtained from proteolytic digestions of the *T. acidophilum* pyruvate kinase, were used to probe for the gene encoding the enzyme. Encouragingly, both probes were found to hybridise to the same regions of *T. acidophilum* genomic DNA - a 7kb EcoRI fragment, a 4.5kb EcoRI/BamHI fragment and a 2kb EcoRI/HinDIII fragment. The 2kb fragment was cloned into pBluescriptII vector and partially sequenced using an M13 universal primer and O2. The parts of the insert sequenced proved to contain no region corresponding to either oligonucleotide sequence. It is possible that, if this insert does indeed contain the *T. acidophilum* pyruvate kinase gene, then the sequences so far obtained from the open reading frames represent the N-terminal region (Fig.6.10) and a non-conserved internal region of the protein (Fig. 6.11) and that the hybridisation sites of the probes have not yet been reached. To complete the sequencing of the insert will require the construction of more oligonucleotide sequencing primers based on the sequence so far obtained. Attempts to sequence the insert using O1 as primer were unsuccessful - possibly the conditions used in the sequencing reactions (see section 2.2.25) were not suitable for successful priming using this oligonucleotide.

AGTAGCCCTGGATACAAAACCCGACCTGCGAACACACGTATCAACCAAGGA

GTCGAGCCAGTGTCGAGCAACATCGATTCGAAC

Corresponds to sequence:

VALDTKPDLRTHVSTKESSQCRATSIRT

Figure 6.16: Partial sequence of 0.6 kb pCRII PCR product insert obtained using M13 universal reverse primer. The polypeptide sequence derived from the open reading frame is shown below the nucleotide sequence. A hybridisation site for probe O1 has been identified (underlined).

6.3.2 Use of probes O1 and O4

Probe O1 corresponds to the sequence of a peptide containing a highly conserved region of the *T. acidophilum* pyruvate kinase. The presence of such a region suggested that construction of probes based on other conserved regions of pyruvate kinases might prove useful. Probe O4 was designed based on this assumption and was used, in conjunction with O1, to produce a PCR product the size and orientation of which was predicted from the pyruvate kinase sequence alignment (Fig. 1.7). This product was cloned into the vector pCRII and partially sequenced using M13 universal forward and reverse primers. Two stretches of sequence were obtained from the 0.6kb insert, each of which contained a region corresponding to one of the PCR primers (Figs. 6.15 and 6.16): AGTAGCCCTGGATACAAAACCCGACCTGCGAACA corresponding to O1 and TACCAGCGTTCCCCGCTAGAGCCGCAGCTCTATGGG corresponding to O4. Only lack of time prevented the determination of the entire sequence of the PCR insert.

The four stretches of sequence obtained from the two clones were translated and analysed by multiple alignment and secondary structure prediction techniques and this is fully described in Chapter 7.

Chapter 7: Analysis of the Partial *T. acidophilum* Pyruvate Kinase Sequence

CHAPTER 7: ANALYSIS OF THE PARTIAL *T.ACIDOPHILUM* PYRUVATE KINASE SEQUENCE

7.1 Introduction

7.1.1 Programs used

The recent merging of the main nucleotide sequence databases, EMBL and GENBANK, has made easily available over 100 million bases of DNA information. This huge resource can be accessed via a number of sequence alignment and secondary structure prediction programs in order to compare newly obtained protein and DNA sequences with previously entered data. Four such programs have been used to analyse the partial sequences obtained from the 2kb and 0.6kb cloned inserts derived from the *T. acidophilum* genome and described in the previous chapter.

The multiple alignment program CLUSTAL (see section 1.1.2) was used to align the two larger stretches of sequence - KLARCSYRWYFLHCAPVLS PPPSWSNSFGVRLLYDSQDRHNSHDYKMSRRSRYYIEDNKNSVATSRNKRGSACE ETVTPRYKVKVPVAPIHTALV and MVARGDLGVEIPLATQEARPGKHSSRSAAG RVTAGNDHLGSSHGSNYVSNSADITVDLSNRISTSIDKWTQFNNRRVLRCVCMLAE AASVKPGVPN - to the alignment of pyruvate kinase sequences stored in the database. Since CLUSTAL is only informative when used for longer sequences, the single alignment program GAP was used for comparing the shorter sequences - AVALDTKPDLRTHVSTKESSQCRATSIRT and SRVTAGNDHQDRAMVRTTYGTHKISQWIYRNRISTSID - to representative pyruvate kinase sequences from human liver, *E. coli* and *S. cerevisiae.* The longer sequences were also aligned using GAP with these pyruvate kinases in order to compare the two alignment programs. In addition, secondary structure predictions were performed using the program PREDICT. This program uses a number of indicators (for example the Chou and Fasman index) to predict the folds of the primary structure entered.

7.1.2 Proteins as phylogenetic markers

The number of archaebacterial protein sequences being used for phylogenetic comparison with the equivalent eukaryotic and

eubacterial proteins is increasing gradually as their gene sequences become available. Although these comparisons have generally confirmed that the archaebacteria fall into a phylogenetic grouping deserving separate classification, some contradictory evidence regarding the interrelationships of the three domains has been gathered. Some archaebacterial proteins, for example glutamine synthetase (Sanangelantoni *et al.*, 1990), show greater sequence similarity to their eubacterial than their eukaryotic counterparts whereas the sequence of archaebacterial elongation factor-Tu (EF-Tu, Lechner and Bock, 1987) is more closely related to its eukaryotic than to its eubacterial counterpart. Furthermore, yet other archaebacterial proteins are very divergent from both their eubacterial and eukaryotic counterparts which closely resemble one another (for example, GAPDH, Hensel *et al.*, 1989). The alignments of the partial *T. acidophilum* pyruvate kinase sequence presented here (the first for an archaebacterial pyruvate kinase) add to this store of information and contribute towards the ultimate goal of a clearer understanding of the true interrelationships between the three domains.

7.1.3 Secondary structure predictions

It is generally the case that the sequences of homologous proteins are less well conserved than are their higher level structures. It may be better, therefore, to consider homology in terms of secondary and tertiary structure as the basis for comparing two proteins. An obvious caveat for this hypothesis, however, is that convergent evolutionary processes are more likely to produce homology of this type than that between two amino acid sequences and may, therefore, indicate a relationship between two proteins which is not in fact real.

In the absence of X-ray data for protein molecules, a number of methods have been developed to predict their secondary structures (α-helices, turns and β-sheets) from their amino acid sequences. These methods, whilst having success rates of only around 65%, can be very useful in providing an indication of potential secondary structural features within a protein. The predictions presented for the archaebacterial sequence fragments presented here were obtained using the program PREDICT. This program predicts structure based on the combination of seven separate methods [Burgess *et al.* (1974), Chou

and Fasman (1974), as modified by Lenstra *et al.* (1977) and with the updated indices of Geisow and Roberts (1980); Dufton and Hider (1977); Garnier *et al.* (1978); Lim (1974); McLachlan (1977); Nagano (1973)] as suggested by Eliopoulos *et al.* (1982). Because the method of McLachlan (1977) is included, the rules for the final predictions are: (a) a residue is considered likely to form a helical or sheet conformation if at least four of the methods predict it to be so; (b) a sheet or helix requires at least four consecutive residues before it is deemed to exist; (c) any regions not predicted to be helical, sheet or turns are considered to be irregular.

7.1.4 Thermostable proteins

Thermophilic proteins exist and function routinely in conditions which would rapidly denature proteins from mesophilic organisms. Many studies have been undertaken to attempt to elucidate the structural factors which stabilise these proteins ranging from comparisons of amino acid compositions to full analysis of three dimensional structures. The latter studies have yielded the best insights into the factors which control the thermal stability of a protein. Perutz and Raidt (1975) concluded that thermal stability is enhanced in ferredoxins and haemoglobins by the presence of additional salt bridges and/or hydrogen bonds. This can be explained by the observation that the stability of a protein can be increased by additional stabilisation energy of only 5-10 kcal (Perutz, 1978) - salt bridges contribute 1-3 kcal each. In other molecules however, more draconian changes may be necessary to confer thermal stability. The GAPDH of lobster is extensively different from that of *B. stearothermophilus* (Walker *et al.*, 1980). In this case it would appear that a number of changes have been made to permit a balance between stability and the flexibility required for activity.

Mutagenesis studies have pinpointed some amino acid replacements which can significantly increase the thermal stability of a protein. The proline theory of Suzuki (1989) has already been mentioned (section 5.3). Other changes, for example arg to his, asn to thr or asn to ile, have been shown to increase thermostability and thermoresistance in protein molecules and may, in future, lead to the elucidation of 'rules' for stabilisation of commercially applicable proteins.

In the cases where the alignment of a stretch of archaebacterial pyruvate kinase sequence to another mesophilic counterpart is a good one (above 30% identity), amino acid substitutions have been logged and these data are presented below.

7.2 Results

7.2.1 Codon usage within the pyruvate kinase sequence stretches

All archaebacterial genes so far analysed have been found to use the standard genetic code. The choice of codons in these genes, like that in the genes of eubacteria and eukaryotes, is dependent on the relative abundance of iso-accepting tRNA species. An implication of this is that the codon preference of organisms is highly selective and, hence, it cannot be used to measure long term evolutionary events across kingdoms (Woese, 1987). It can be used, however, to measure close evolutionary events given that individual organisms and those which are closely related to one another show similar codon preferences.

The codon usage for the partial *T. acidophilum* pyruvate kinase gene is shown in Table 7.1. This may be compared with the codon usage table for the *T. acidophilum* citrate synthase gene shown in Table 6.1.

7.2.2 Alignments of the partial sequences

The partial sequence stretches from the *T. acidophilum* pyruvate kinase were compared with those from eukaryotic and eubacterial species by two computer-aided sequence alignment methods - CLUSTAL and GAP. Fig. 7.1 shows a CLUSTAL alignment of the sequence shown in Fig. 6.15 and Fig. 7.2 shows a CLUSTAL alignment of the sequence shown in Fig. 6.10. Figs. 7.3 - 7.6 show GAP alignments of each of the four sequences obtained with pyruvate kinase sequences obtained from human liver, yeast (*Saccharomyces cerevisiae*) and from a newly obtained *E. coli* enzyme fragment. Sequence identities and similarities between each sequence pair have been calculated for each alignment and are tabulated in Table 7.2.

F	TTT	2	**S**	TCT	3	**Y**	TAT	9	**C**	TGT	2
F	TTC	3	**S**	TCC	5	**Y**	TAC	4	**C**	TGC	3
L	TTA	4	**S**	TCA	4	*	TAA	-	*	TGA	-
L	TTG	5	**S**	TCG	3	*	TAG	-	**W**	TGG	6
L	CTT	3	**P**	CCT	5	**H**	CAT	6	**R**	CGT	2
L	CTC	1	**P**	CCC	4	**H**	CAC	6	**R**	CGC	3
L	CTA	5	**P**	CCA	4	**Q**	CAA	5	**R**	CGA	6
L	CTG	3	**P**	CCG	5	**Q**	CAG	3	**R**	CGG	5
I	ATT	4	**T**	ACT	4	**N**	AAT	8	**S**	AGT	10
I	ATC	1	**T**	ACC	1	**N**	AAC	3	**S**	AGC	7
I	ATA	3	**T**	ACA	4	**K**	AAA	5	**R**	AGA	1
M	ATG	2	**T**	ACG	4	**K**	AAG	4	**R**	AGG	7
V	GTT	3	**A**	GCT	2	**D**	GAT	5	**G**	GGT	1
V	GTC	4	**A**	GCC	8	**D**	GAC	5	**G**	GGC	2
V	GTA	7	**A**	GCA	3	**E**	GAA	5	**G**	GGA	1
V	GTG	5	**A**	GCG	2	**E**	GAG	3	**G**	GGG	3

Table 7.1: Codon usage of the *T. acidophilum* pyruvate kinase partial sequence.

```
 Gap fixed =  10 Gap vary. =  10
 * :=>  match across all seqs.
 . :=>  conservative substitutions

ratmus1          MVARGDLGIEIPAEKVFLAQKMMIGR--CNRAGKPVICATQMLESMIKKPRPTRAEGSDV
ratmus1          MVARGDLGIEIPAEKVFLAQKMMIGR--CNRAGKPVICATQMLESMIKKPRPTRAEGSDV
catmus1          MVARGDLGIEIPAEKVFLAQKMMIGR--CNRAGKPVICATQMLESMIKKPRPTRAEGSDV
chimus           MVARGDLGIEIPAEKVFLAQKMMIGR--CNRAGKPIICATQMLESMIKKPRPTRAEGSDV
humliv           MVARGDLGIEIPAEKVFLAQKMMIGR--CNLAGKPVVCATQMLESMITKPRPTRAETSDV
ratliv           MVARGDLGIEIPAEKVFLAQKMMIGR--CNLAGKPVVCATQMLESMITKARPTRAETSDV
ratrbc           MVARGDLGIEIPAEKVFLAQKMMIGR--CNLAGKPVVCATQMLESMITKARPTRAETSDV
potCy            MVARGDLGMEIPVEKIFLAQKMMIYK--CNLAGKAVVTATQMLESMIKSPAPTRAEATDV
Anid             MVARGDLGIEIPAPKVFIAQKMMIAK--CNIKGKPVICATQMLESMTYNPRPTRAEVSDV
yea              MVARGDLGIEIPAPEVLAVQKKLIAK--SNLAGKPVICATQMLESMTYNPRPTRAEVSDV
Bst              MVARGDLGVEIPAEEVPLIQKLLIKK--CNMLGKPVITATQMLDSMQRNPRPTRAEASDV
Tacil            MVARGDLGVEIP----LATQEARPGKHSSRSAAGRVTAGNDHLGS--SHGSNYVSNSADI
                 ********.***       *.    . .. . .. ... *.*          .. .*.

ratmus1          ANAVLDGADCIMLSGETAKGDYPLEAVRM--QHLIAREAEAA
ratmus1          ANAVLDGADCIMLSGETAKGDYPLEAVRM--QHLIAREAEA-
catmus1          ANAVLDGADCIMLSGETAKGDYPLEAVRM--QHLIAREAEA-
chimus           ANAVLDGADCIMLSGETAKGDYPLEAVRM--QHAIAREA---
humliv           ANAVLDGADCIMLSGETAKGNFPVEAVKM--QHRIAREAEAA
ratliv           ANAVLDGADCIMLSGETAKGSFPVEAVMM--QHAIAREAEAA
ratrbc           ANAVLDGADCIMLSGETAKGNFPVEAVMM--QHAIAREAEAA
potCy            ANAVLDGTDCVMLSGESAAGAYPELAVKIMSRICI----EAE
Anid             ANAVLDGADCVMLSGETAKGNYPCEAVTMMSETCLLAE----
yea              GNAILDGADCVMLSGETAKGNYPINAVTTMAETAVIAEQAIA
Bst              ANAIFDGTDAVMLSGETAAGQYPVEAVKTMHQIALRTE----
Tacil            TVDLSNRISTSIDKWTQFNNRRVLRCVCMLAEAASVKPGVPN
                 . .. . . . . .  .      *
```

Figure 7.1: CLUSTAL alignment of conserved archaebacterial pyruvate kinase peptide MVARG with pyruvate kinase complete sequences.

```
 Gap fixed =  10 Gap vary. =  10
 * :=>  match across all seqs.
 . :=>  conservative substitutions

catmus1          SKPHSDVGTAFIQTQQLHAAMADTFLEHMCRLDIDSPPITA-RNTGIICTIGPASRSVEI
chimus           SKHH-DAGTAFIQTQQLHAAMADTFLEHMCRLDIDSEPTIA-RNTGIICTIGPASRSVDK
potCy            ----------MANIDIAGIMKD--LPNDGRI----------PKTKIVCTLGPSSRTVPM
Anid             AASSS--------LDHLSNRMK---LEWHSKLNTEMVPAKNFRRTSIICTIGPKTNSVEK
yea              S----------------R---LERLTSLN--VVAGSDLRRTSIIGTIGPKTNNPET
Bst              K----------------------------------------RKTKIVCTIGPASESVDK
                 KLARCSYRWYFLHCAPVLSPPP----SWSNSFGVRLLYDSQDRHNSHDYKMSRRSRY---
                                                          ...    .... .

catmus1          LKEMIKSGMNVARLNFSHGTHEYHAETIKNVRAATESFASDPIRYRPVAV
chimus           LKEMIKSGMNVARLNFSHGTHEYHEGTIKNVREATESFASDPITYRPV--
potCy            LEKLLRAGMNVARFNFSHGTHEYGQETLDNLKIAMQN---------TQIL
Anid             INALRRAGLNVVRMNFSHGSYEYHQSVIDHAREAEKQAAGRPV-------
yea              LVALRKAGLNIVRMNFSHGSYEYHKSVIDNARKSEELYPGRPLA------
Bst              LVQLMEAGMNVARLNFSHGDHEEHGRRIANIREAAK---------RTGRT
                 YIEDNKNSVATSRNK--RGS-ACEETVTPRYKVKV---PVAPI--HTALV
                       ..... * .  .*. .      . .
```

Figure 7.2: CLUSTAL alignment of N-terminal archaebacterial pyruvate kinase peptide KLARC with pyruvate kinase complete sequences.

GAP alignments:

sequence	source	%identity	%similarity
MVARG...	huliv	27.9	51.2
	ysc	34.8	56.2
	eco	29.9	51.7
AVALD...	huliv	32.1	67.9
	ysc	34.5	51.7
	eco	20.7*	34.5*
SRVTA...	huliv	19.4	36.1
	ysc	21.1	44.7
	eco	31.6	50.0
KLARC...	huliv	14.3	42.9
	ysc	28.9	48.9
	eco	19.8	45.1

Table 7.2: Summary of GAP alignments of *T. acidophilum* pyruvate kinase sequence stretches and complete pyruvate kinase sequences from human liver (huliv), *Saccharomyces cerevisiae* (ysc) and a fragment from a new *E. coli* sequence. * - the incomplete *E. coli* sequence does not possess the conserved region identified in the other two sequences by this region of the *T. acidophilum* enzyme. The GAP program has, therefore, aligned the stretch inappropriately and the % identity and similarity figures consequently suffer.

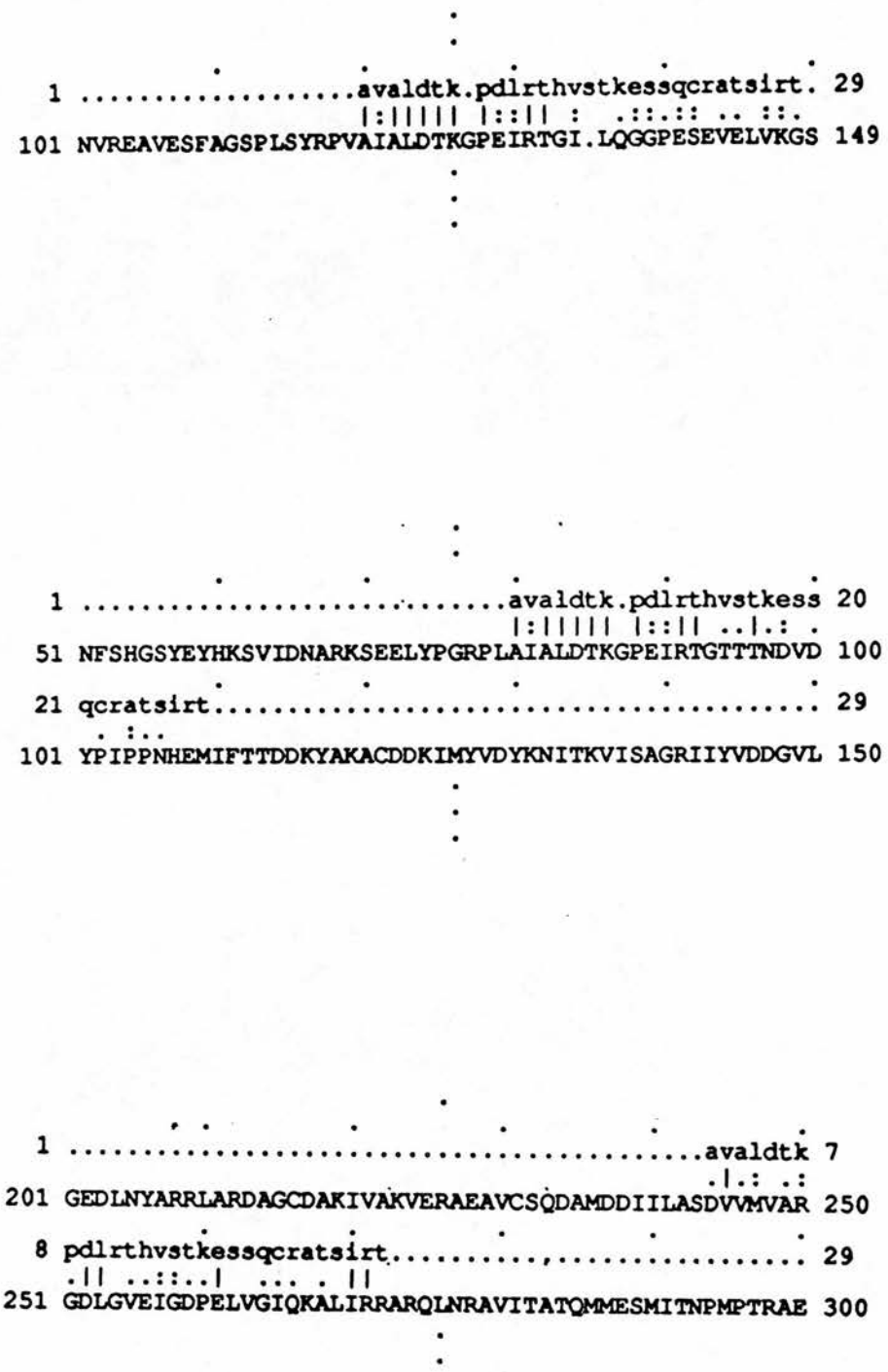
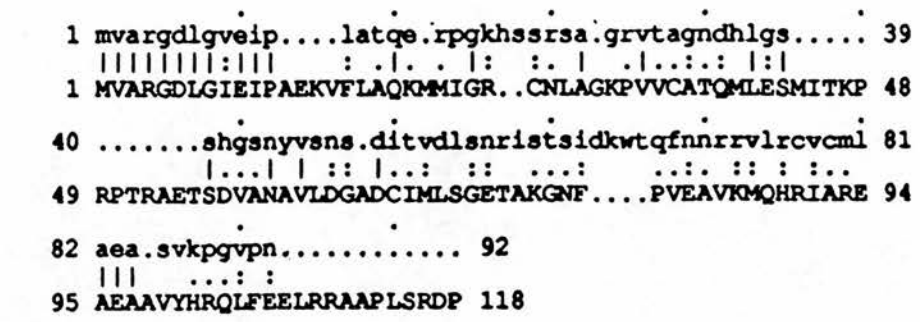
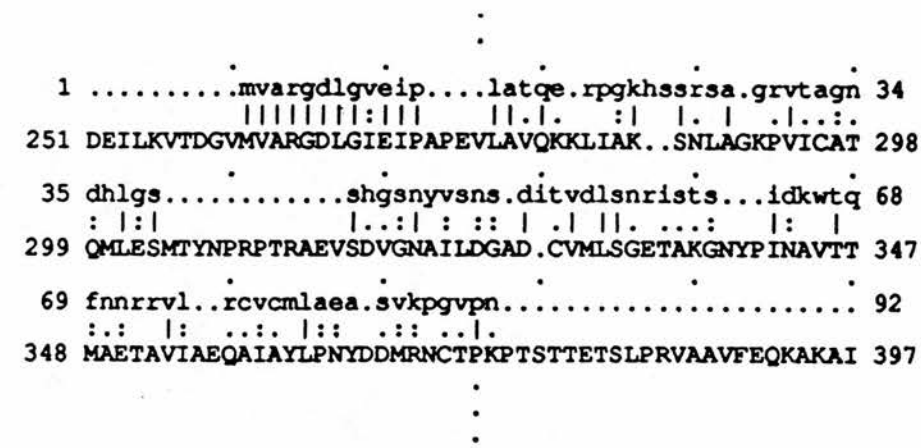

1 .....avaldtk.pdlrthvstkessqcratsirt. 29
101 NVREAVESFAGSPLSYRPVAIALDTKGPEIRTGI.LQGGPESEVELVKGS 149
1 .....avaldtk.pdlrthvstkess 20
51 NFSHGSYEYHKSVIDNARKSEELYPGRPLAIALDTKGPEIRTGTTTNDVD 100
21 qcratsirt..... 29
101 YPIPPNHEMIFTTDDKYAKACDDKIMYVDYKNITKVISAGRIIYVDDGVL 150
1 .....avaldtk 7
201 GEDLNYARRLARDAGCDAKIVAKVERAEAVCSQDAMDDIILASDVVMVAR 250
8 pdlrthvstkessqcratsirt..... 29
251 GDLGVEIGDPELVGIQKALIRRARQLNRAVITATQMMESMITNPMPTRAE 300

Figure 7.3: GAP alignments of peptide AVALD with pyruvate kinase sequences of human liver (top), *S. cerevisiae* (middle) and *E. coli* (bottom). Percent identity and similarity values were respectively: huliv, 32.1%, 67.9%; ysc, 34.5%, 51.7%; eco, 20.7%, 34.5%.


1 mvargdlgveip....latqe.rpgkhssrsa.grvtagndhlgs..... 39
1 MVARGDLGIEIPAEKVFLAQKMMIGR..CNLAGKPVVCATQMLESMITKP 48
40 .....shgsnyvsns.ditvdlsnristsidkwtqfnnrrvlrcvcml 81
49 RPTRAETSDVANAVLDGADCIMLSGETAKGNF....PVEAVKMQHRIARE 94
82 aea.svkpgvpn..... 92
95 AEAAVYHRQLFEELRRAAPLSRDP 118


1 .....mvargdlgveip....latqe.rpgkhssrsa.grvtagn 34
251 DEILKVTDGVMVARGDLGIEIPAPEVLAVQKKLIAK..SNLAGKPVICAT 298
35 dhlgs.....shgsnyvsns.ditvdlsnrists...idkwtq 68
299 QMLESMTYNPRPTRAEVSDVGNAILDGAD.CVMLSGETAKGNYPINAVTT 347
69 fnnrrvl..rcvcmlaea.svkpgvpn..... 92
348 MAETAVIAEQAIAYLPNYDDMRNCTPKPTSTTETSLPRVAAVFEQKAKAI 397

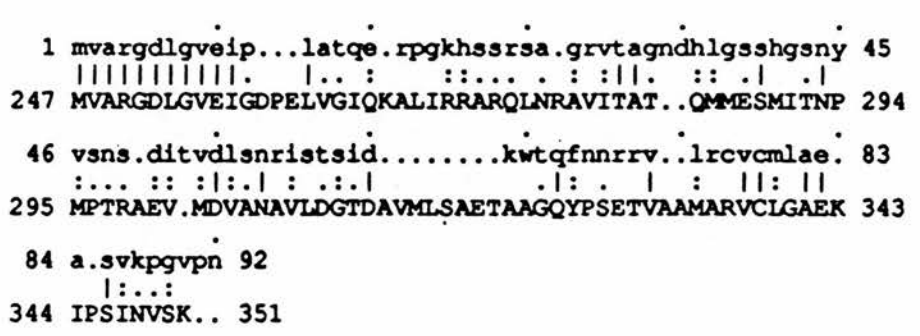
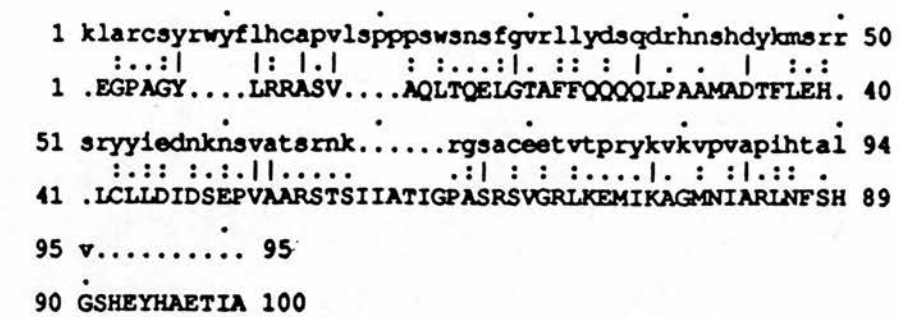
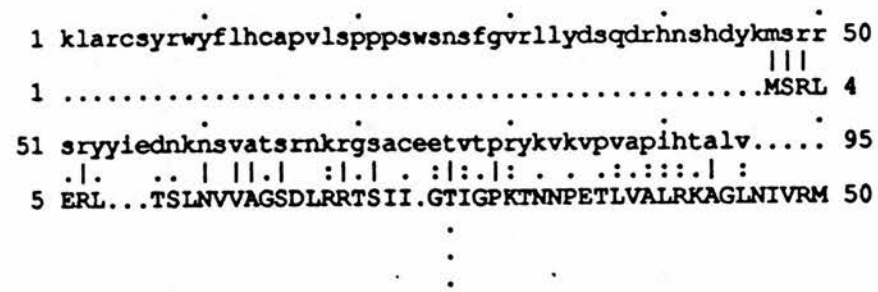

1 mvargdlgveip...latqe.rpgkhssrsa.grvtagndhlgsshgsny 45
247 MVARGDLGVEIGDPELVGIQKALIRRARQLNRAVITAT..QMMESMITNP 294
46 vsns.ditvdlsnristsid.....kwtqfnnrrv..lrcvcmlae. 83
295 MPTRAEV.MDVANAVLDGTDAVMLSAETAAGQYPSETVAAMARVCLGAEK 343
84 a.svkpgvpn 92
344 IPSINVSK.. 351

Figure 7.4: GAP alignments of peptide MVARG with pyruvate kinase sequences of human liver (top), *S. cerevisiae* (middle) and *E. coli* (bottom). Percent identity and similarity values were respectively: huliv, 27.9%, 51.2%; ysc, 34.8%, 56.2%; eco, 29.9%, 51.7%.


1 klarcsyrwyflhcapvlspppswsnsfgvrllydsqdrhnshdykmsrr 50
1 .EGPAGY....LRRASV....AQLTQELGTAFFQQQQLPAAMADTFLEH. 40
51 sryyiednknsvatsrnk.....rgsaceetvtprykvkvpvapihtal 94
41 .LCLLDIDSEPVAARSTSIIATIGPASRSVGRLKEMIKAGMNIARLNFSH 89
95 v..... 95
90 GSHEYHAETIA 100


1 klarcsyrwyflhcapvlspppswsnsfgvrllydsqdrhnshdykmsrr 50
1 .....MSRL 4
51 sryyiednknsvatsrnkrgsaceetvtprykvkvpvapihtalv..... 95
5 ERL...TSLNVVAGSDLRRTSII.GTIGPKTNNPETLVALRKAGLNIVRM 50

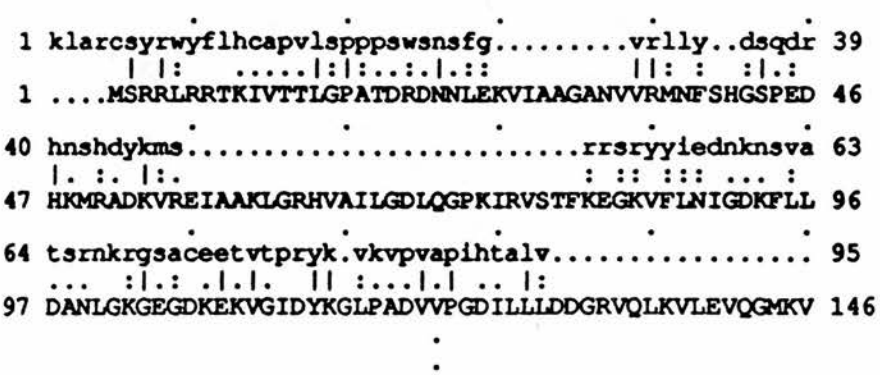

1 klarcsyrwyflhcapvlspppswsnsfg.....vrlly..dsqdr 39
1 ....MSRRLRRTKIVTTLGPATDRDNNLEKVIAAGANVVRMNFSHGSPED 46
40 hnshdykms.....rrsryyiednknsva 63
47 HKMRADKVREIAAKLGRHVAILGDLQGPKIRVSTFKEGKVFLNIGDKFLL 96
64 tsrnkrgsaceetvtpryk.vkvpvapihtalv..... 95
97 DANLGKGEGDKEKVGIDYKGLPADVVPGDILLLDDGRVQLKVLEVQGMKV 146

Figure 7.5: GAP alignments of peptide KLARC with pyruvate kinase sequences of human liver (top), *S. cerevisiae* (middle) and *E. coli* (bottom). Percent identity and similarity values were respectively: huliv, 14.3%, 42.9%; ysc, 28.9%, 48.9%; eco, 19.8%, 45.1%.

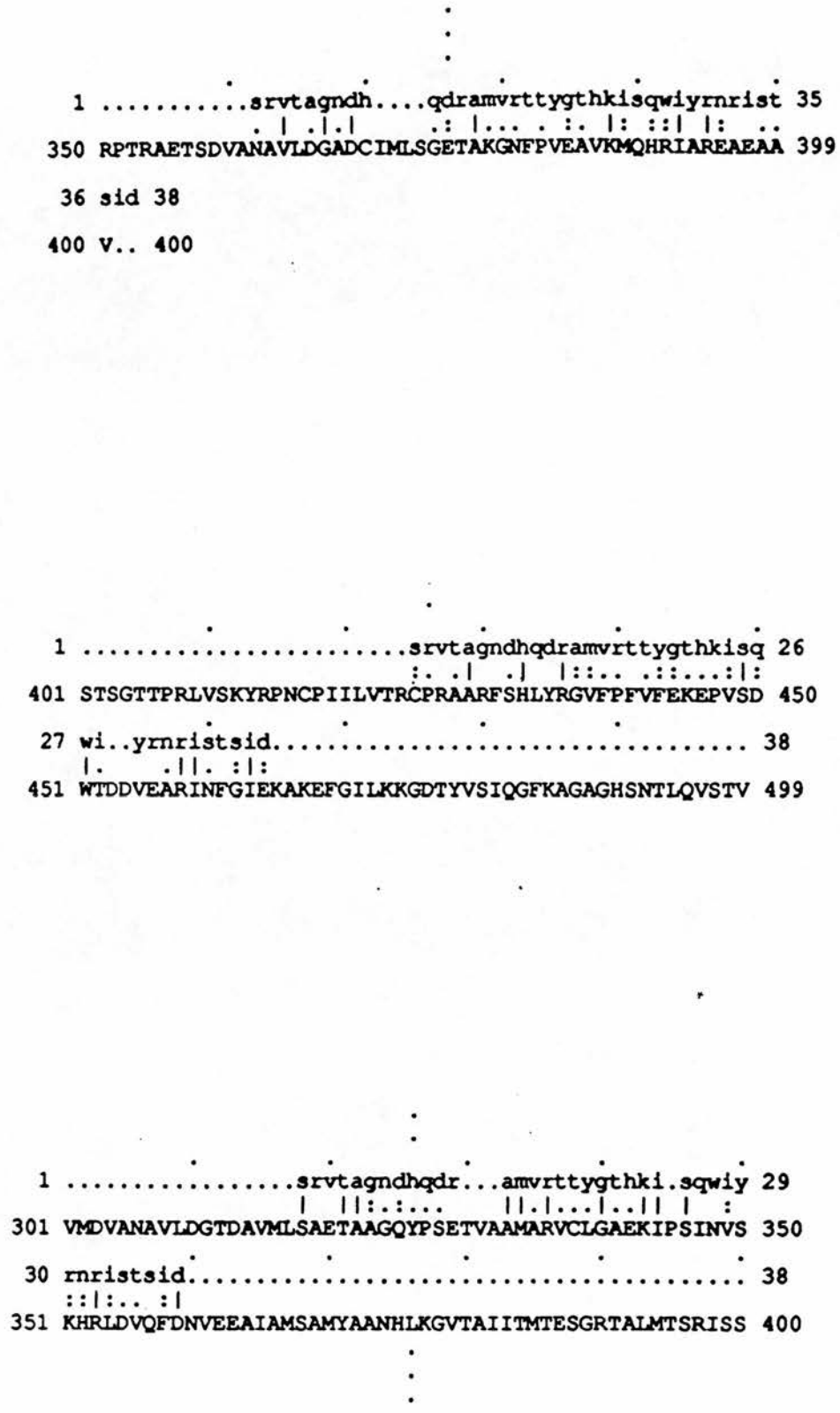
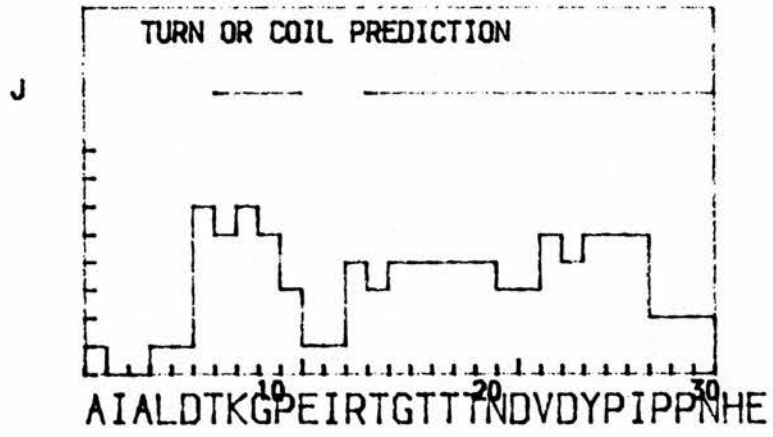

1 .....srvtagndh....qdramvrttygthkisqwiyrnrist 35
350 RPTRAETSDVANAVLDGADCIMLSGETAKGNFPVEAVKMQHRIAREAEAA 399
36 sid 38
400 V.. 400
1 .....srvtagndhqdramvrttygthkisq 26
401 STSGTTPRLVSKYRPNCPIILVTRCPRAARFSHLYRGVFPFVFEKEPVSD 450
27 wi..yrnristsid..... 38
451 WTDDVEARINFGIEKAKEFGILKKGDTYVSIQGFKAGAGHSNTLQVSTV 499
1 .....srvtagndhqdr...amvrttygthki.sqwiy 29
301 VMDVANAVLDGTDAVMLSAETAAGQYPSETVAAMARVCLGAEKIPSINVS 350
30 rnristsid..... 38
351 KHRLDVQFDNVEEAIAMSAMYAANHLKGVTAIITMTESGRTALMTSRISS 400

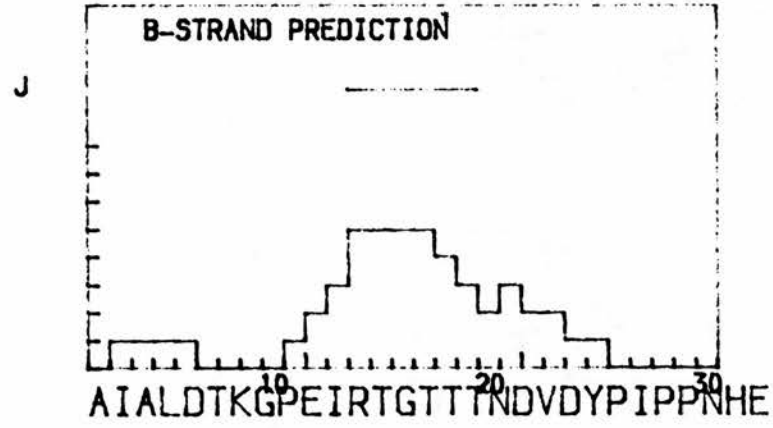
Figure 7.6: GAP alignments of peptide SVRTA with pyruvate kinase sequences of human liver (top), *S. cerevisiae* (middle), *E. coli* (bottom). Percent identity and similarity values were respectively: huliv, 19.4%, 36.1%; ysc, 21.1%, 44.7%; eco, 31.6%, 50.0%.

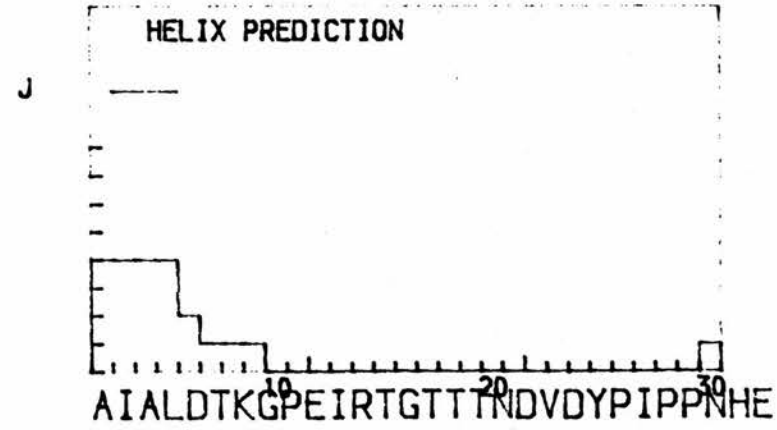
```
            hhhhtttt--bbbbbbtttttbbbbbbbb
archae      AVALDTKPDLRTHVSTKESSQCRATSIRT

            hhhhttttttbbbbbbbbbttttttttt
yeast       AIALDTKGPEIRTGTTTNDVDYPIPPNHE
```

Figure 7.7: Secondary structure prediction for the conserved archaebacterial pyruvate kinase sequence AVALD (top) and the corresponding region of the yeast (*S. cerevisiae*) sequence identified by the GAP alignment program (bottom). h = alpha helix, b = beta sheet, t = turn or coil.

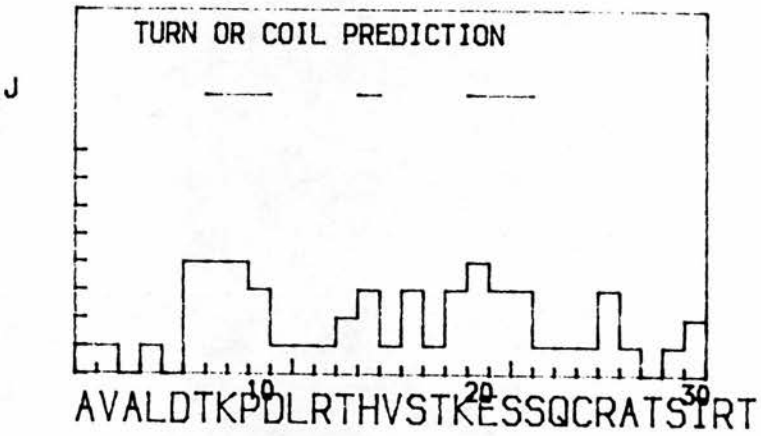
YEAS

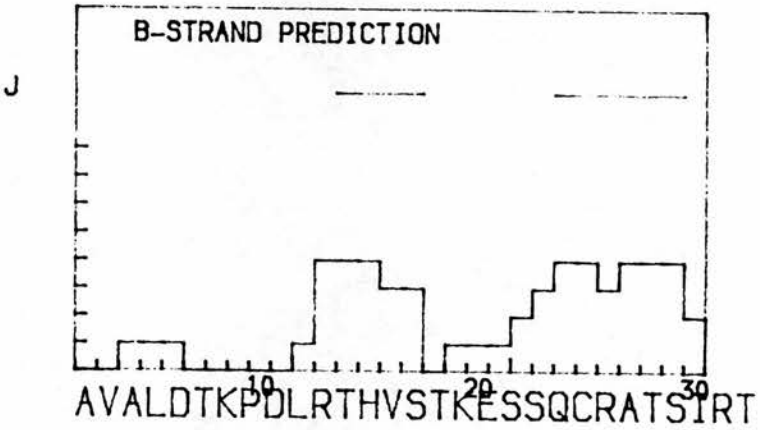

TURN OR COIL PREDICTION
J
10
20
30
AIALDTKGPEIRTGTTTNDVDYPIPPNHE

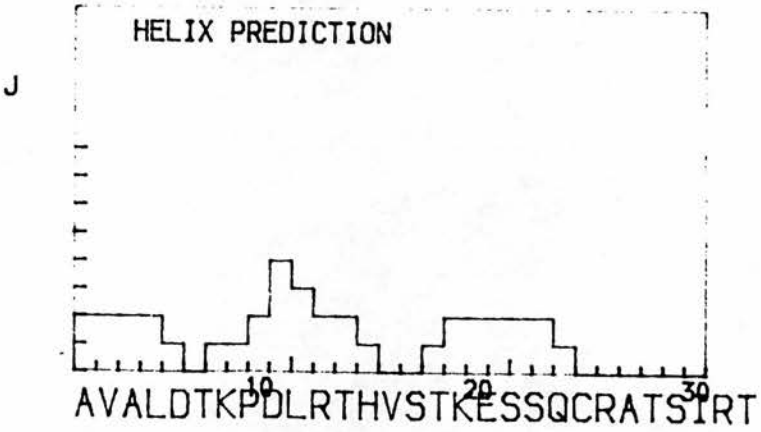

B-STRAND PREDICTION
J
10
20
30
AIALDTKGPEIRTGTTTNDVDYPIPPNHE


HELIX PREDICTION
J
10
20
30
AIALDTKGPEIRTGTTTNDVDYPIPPNHE

PRED

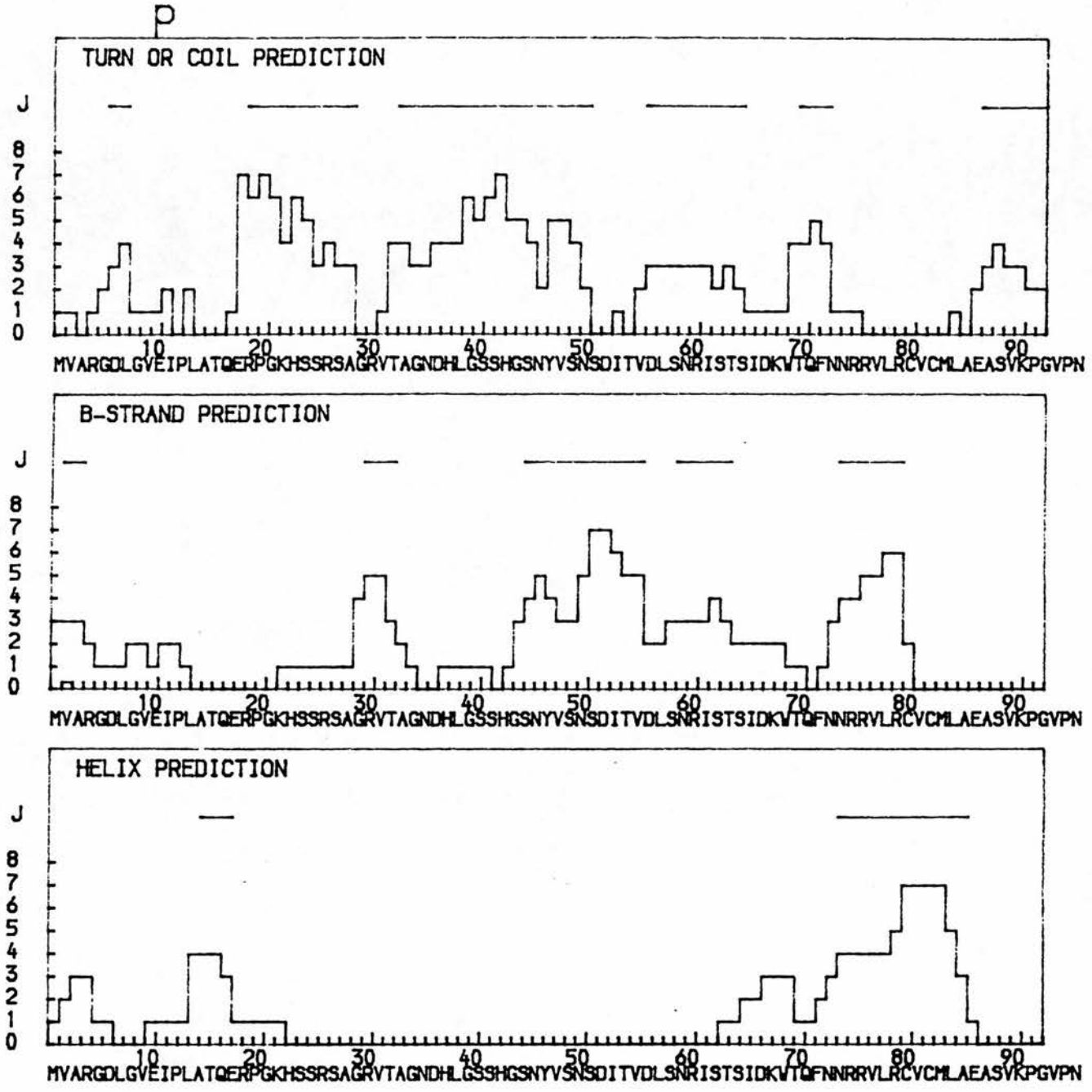

TURN OR COIL PREDICTION
J
10
20
30
AVALDTKPDLRTHVSTKESSQCRATSIRT


B-STRAND PREDICTION
J
10
20
30
AVALDTKPDLRTHVSTKESSQCRATSIRT

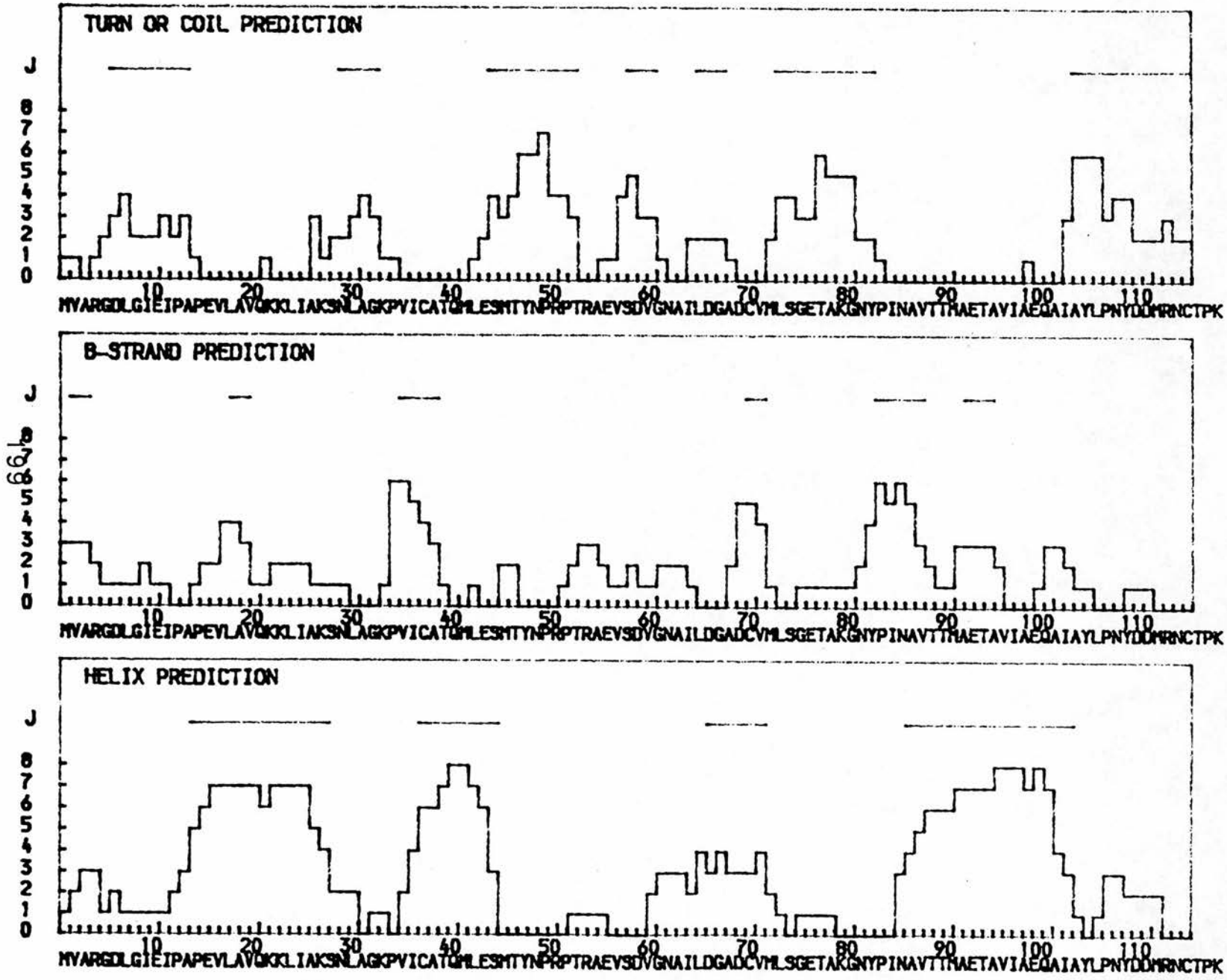

HELIX PREDICTION
J
10
20
30
AVALDTKPDLRTHVSTKESSQCRATSIRT

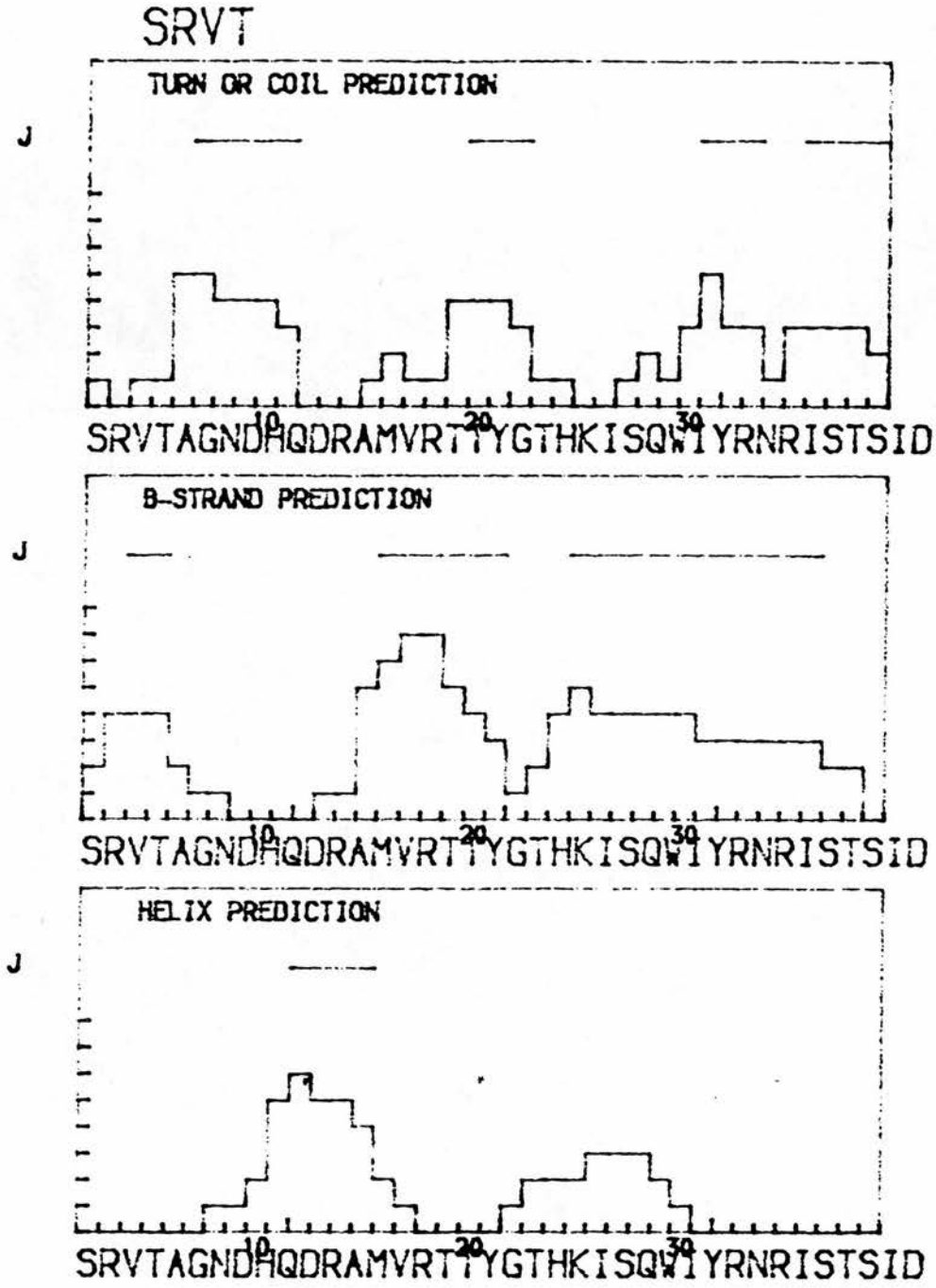
```
                bbbbtttt----hhhhhttttttttttbbbbtttttttttb
archae          MVARGDLGVEIPLATQERPGKHSSRSAGRVTAGNDHLGSSHGS

                bbbbbbbbbbbbbtttbbbb----ttthhhhhhhhhhhhhht
                NYVSNSDITVDLSNRISTSIDKWTQFNNRRVLRCVCMLAEASV

                tttttt
                KPGVPN


                bbbbtttttttthhhhhhhhhhhhhhttttt---hhhhhhhht
yeast           MVARGDLGIEIPAPEVLAVQKKLIAKSNLAGKPVICATQMLESMTY

                ttttttbbbbtttt---ttttbbbbbtttttttttbbbbhhhh
                PRPTRAEVSDVGNAILDGADCVMLSGETAKGNYPINAVTTMAETAV

                hhhhhhhhhhhttttttttt
                IAEQAIAYLPNYDDMRNCTPK
```

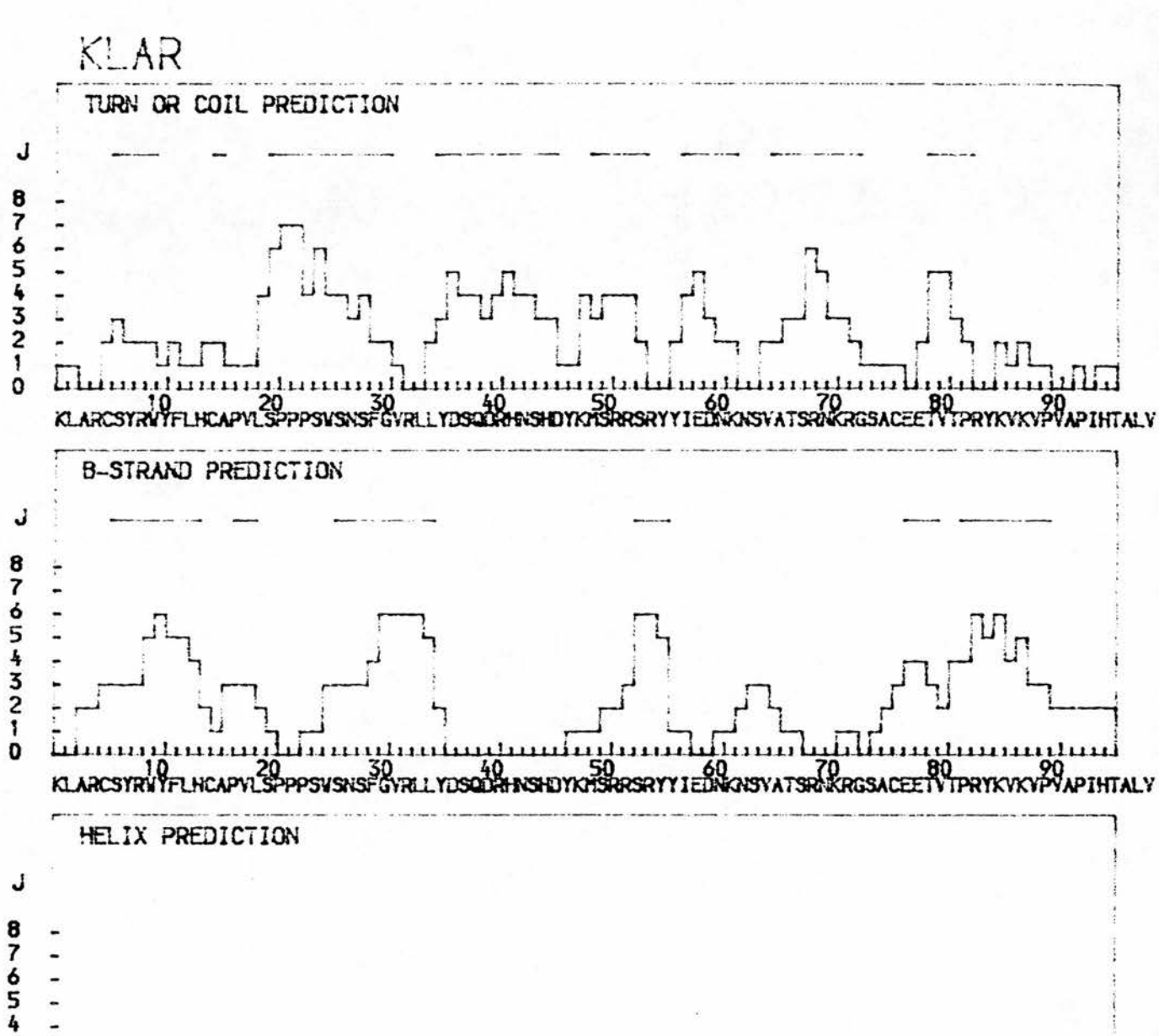
Figure 7.8: Secondary structure predictions for the conserved archaebacterial pyruvate kinase sequence fragment MVARG (top) and the corresponding region of the yeast (*S. cerevisiae*) enzyme sequence identified by the GAP alignment program (bottom). h = alpha helix, b = beta sheet, t = turn or coil. p.198, 199 show prediction outputs for archaebacterial and yeast sequences respectively.


P
TURN OR COIL PREDICTION
J
8
7
6
5
4
3
2
1
0
10
20
30
40
50
60
70
80
90
MVARGDLGVEIPLATQERPGKHSSRSAGRVTAGNDHLGSSHGSNYVSNSDITVDLSNRISTSIDKVTQFNNRRVLRCVCMLAEASVKPGVPN
B-STRAND PREDICTION
J
8
7
6
5
4
3
2
1
0
10
20
30
40
50
60
70
80
90
MVARGDLGVEIPLATQERPGKHSSRSAGRVTAGNDHLGSSHGSNYVSNSDITVDLSNRISTSIDKVTQFNNRRVLRCVCMLAEASVKPGVPN
HELIX PREDICTION
J
8
7
6
5
4
3
2
1
0
10
20
30
40
50
60
70
80
90
MVARGDLGVEIPLATQERPGKHSSRSAGRVTAGNDHLGSSHGSNYVSNSDITVDLSNRISTSIDKVTQFNNRRVLRCVCMLAEASVKPGVPN

YEAS


TURN OR COIL PREDICTION
J
0 1 2 3 4 5 6 7 8
10 20 30 40 50 60 70 80 90 100 110
MVARGDLGIEIPAPEVLAVGKKLIAKSNLAGKPVICATQMLESMTYNPRPTRAEVSDVGNAILDGADCVMLSGETAKGNYPINAVTTMAETAVIAEQAIAYLPNYDDMRNCTPK
B-STRAND PREDICTION
J
0 1 2 3 4 5 6 7 8
10 20 30 40 50 60 70 80 90 100 110
MVARGDLGIEIPAPEVLAVGKKLIAKSNLAGKPVICATQMLESMTYNPRPTRAEVSDVGNAILDGADCVMLSGETAKGNYPINAVTTMAETAVIAEQAIAYLPNYDDMRNCTPK
HELIX PREDICTION
J
0 1 2 3 4 5 6 7 8
10 20 30 40 50 60 70 80 90 100 110
MVARGDLGIEIPAPEVLAVGKKLIAKSNLAGKPVICATQMLESMTYNPRPTRAEVSDVGNAILDGADCVMLSGETAKGNYPINAVTTMAETAVIAEQAIAYLPNYDDMRNCTPK


SRVT
TURN OR COIL PREDICTION
J
SRVTAGNDHQDRAMVRTYGTHKISQWIYRNRISTSID
10
20
30
B-STRAND PREDICTION
J
SRVTAGNDHQDRAMVRTYGTHKISQWIYRNRISTSID
10
20
30
HELIX PREDICTION
J
SRVTAGNDHQDRAMVRTYGTHKISQWIYRNRISTSID
10
20
30


KLAR
TURN OR COIL PREDICTION
J
8
7
6
5
4
3
2
1
0
10
20
30
40
50
60
70
80
90
KLARCSYRVYFLHCAPVLSPPPSVSNSFGVRLLYDSQDRHNSHDYKMSRRSRYYIEDNKNSVATSRNKRGSACEETVTPRYKVKVPVAPIHTALV
B-STRAND PREDICTION
J
8
7
6
5
4
3
2
1
0
10
20
30
40
50
60
70
80
90
KLARCSYRVYFLHCAPVLSPPPSVSNSFGVRLLYDSQDRHNSHDYKMSRRSRYYIEDNKNSVATSRNKRGSACEETVTPRYKVKVPVAPIHTALV
HELIX PREDICTION
J
8
7
6
5
4
3
2
1
0
10
20
30
40
50
60
70
80
90
KLARCSYRVYFLHCAPVLSPPPSVSNSFGVRLLYDSQDRHNSHDYKMSRRSRYYIEDNKNSVATSRNKRGSACEETVTPRYKVKVPVAPIHTALV

7.2.3 Secondary structure predictions for the partial sequences

The sequence stretches obtained from the *T. acidophilum* pyruvate kinase were also subjected to secondary structure prediction analysis using the program PREDICT. Figs. 7.7 - 7.10 show the secondary structure predictions obtained for each archaebacterial pyruvate kinase fragment. The histograms show the results of the joint prediction analysis and the lines indicate the regions of likely secondary structure. These regions have been imposed over the sequences of the peptides (note: regions of unassignable structure have been indicated with '-'s). In the two cases where conserved regions have helped pinpoint the position of the alignments, the corresponding regions of yeast pyruvate kinase have also been subjected to the analysis.

7.3 Discussion

7.3.1 Codon usage of the partial pyruvate kinase sequence

Codon usages are often found to vary between genes isolated from the same organism. This would appear to be the case for the pyruvate kinase partial sequence and the citrate synthase complete sequence from *T. acidophilum*. The differences are likely to be a function of the degree of expression of the genes - in *E. coli* for example, strongly expressed genes use codons which correspond to iso-accepting tRNAs which are present in high concentrations whereas more weakly expressed genes utilise these particular codons less frequently (Winnacker, 1987). If this is the reason for the differences observed between the two archaebacterial genes noted here then it would follow that they would not be expressed in the organism to the same extent. This is, in fact, the case as the pyruvate kinase of *T. acidophilum* is typically present in twice the concentration of the citrate synthase of that organism (Smith *et al.*, 1987).

7.3.2 Partial sequence alignments 1 - GAP alignments

GAP alignments were performed for each of the sequences obtained from the archaebacterial pyruvate kinase with complete sequences for the corresponding enzymes isolated from human liver and yeast (*S. cerevisiae*) and a partial *E. coli* sequence. A summary table showing the

percentage identity and similarity between each of the aligned sequences is presented in Table. 7.2. In this table, each of the archaebacterial enzyme sequence stretches is identified by its first five amino acids.

7.3.2.1 Alignments of MVARG, AVALD and KLARC

Three of the peptides - MVARG (96 amino acids), AVALD (29 amino acids) and KLARC (95 amino acids) - show higher percent identity towards the yeast sequence than to either of the others. The values range from 28.9% to 34.8%. These values are significantly above the threshold value for the program (which is between 20 - 25% identity) and the percent similarity values obtained (48.9 - 56.2%) are even more convincing of homology between the archaebacterial and the mesophilic enzymes. The former two peptides contain sequences which are recognisable as aligning with regions found to be conserved throughout all pyruvate kinases isolated from both eukaryotic and eubacterial sources. This is helpful in determining the accuracy of the alignment program - in every case except one, the program accurately aligned the peptides with the corresponding conserved region of the pyruvate kinase sequence. The only exception was that of comparing the peptide AVALD with the incomplete *E. coli* sequence. This is because the bacterial fragment does not contain the conserved region corresponding to the peptide. This alignment result was included here as it gives a good measure of the percentage identity and similarity values expected when two sequences are incorrectly (randomly) matched (20.7% and 34.5% respectively). [It is worth noting at this point that the decision made to assume that the archaebacterial pyruvate kinase would possess the conserved regions in their entirety was invalid. Eubacterial and eukaryotic enzymes all possess the conserved region AVALDTKGPEIRT. The archaebacterial enzyme, however, contains the corresponding sequence AVALDTKPDLRT i.e. the region is 92% similar but only 69% identical]. The alignment of the peptide KLARC to the other sequences proved to be far more problematic. The GAP program in each case identified the peptide as corresponding to the N-termini of the pyruvate kinase sequences although the percent identity values obtained are actually below the threshold value for the program. This is probably because the N-terminal regions of pyruvate

kinases are extremely variable in length and sequence composition (see Fig.1.7).

7.3.2.2 Alignment of SRVTA

This peptide (38 amino acids) was found to align most closely with the bacterial (*E. coli*) sequence displaying 31.6% identity and 50.0% similarity. The values obtained for the alignments with the other, eukaryotic, sequences were much lower, being below the threshold value. This suggests that the alignments of this peptide are not particularly good ones and that it would, therefore, be unwise to extrapolate any conclusions from them.

7.3.3 Partial sequence alignments 2 - CLUSTAL alignments

Given the fact that CLUSTAL alignments are only informative for longer stretches of sequence, only two peptides were subjected to analysis using this program: MVARG and KLARC (corresponding to 96 and 95 amino acids respectively). The alignments are shown in Figs. 7.1 and 7.2. The percentage identity and similarity values obtained from these alignments are not conspicuously good but this may be more a function of the program used than of the sequences themselves: CLUSTAL only works well with sequences that are of equal (or nearly equal) length. If peptides are being compared to full sequences, therefore, the program is not ideal as the complete sequences must be severly edited. With the two peptides used here it proved necessary to edit the complete sequences down to the regions identified by the GAP alignments (if, as with GAP, the program was used to align the peptides to the complete sequences it produced percent identity scores of up to 60% whilst introducing huge gaps into the sequences - some exceeding 200 amino acids in length). This manual editing obviously restricts the program's capability to produce maximum identity scores between the sequences and hence, the alignments have proven to be poor.

7.3.4 Amino acid exchanges

Amino acid exchanges were logged for the alignments obtained using the GAP program for the archaebacterial peptides with the yeast enzyme. These alignments were chosen as they seemed to be the most

reliable statistically. Several points of interest were noted. First, 7 extra proline residues were present in the archaebacterial sequences, 5 in the less well conserved peptides (KLARC and SRVTA). This would correspond well with the proline theory of protein thermostabilisation (Suzuki, 1989) to which reference has already been made. Secondly, the least commonly exchanged residues were histidines and phenylalanines. This is quite reasonable as histidines are often catalytically and structurally important to enzymes and phenylalanines are very bulky moeities. Finally, where there were changes between the mesophilic and thermophilic sequences, the residues most commonly changed to in the archaebacterial peptides were serines, valines, glycines, asparagines, threonines and arginines. Many of these changes can be explained using the 'rules' outlined in Argos *et al.* (1979). For example, the increase in glycines tends to afford greater strength to alpha-helices within a protein. It must be stressed, however, that drawing any conclusions of this type from less than half of the protein sequence is inherently risky and uninformative.

7.3.5 Secondary structure predictions

Secondary structure predictions for each peptide are presented in Figs. 7.7 - 7.10. Of greatest interest are the predicted structures of the two conserved peptides, AVALD and MVARG, which display a great deal of similarity to the corresponding regions of the yeast enzyme (with which they were aligned by the GAP program). The structure of AVALD is: helix-turn-sheet-turn-sheet, whereas the structure of the yeast enzyme in that area is helix-turn-sheet-turn. The structure of the larger MVARG fragment is: sheet-turn-helix-turn-sheet-turn-sheet-turn-sheet-turn-helix-turn. The corresponding yeast enzyme region is: sheet-turn-helix-turn-helix-turn-sheet-turn-sheet-turn-sheet-helix-turn. These similar structures are presumably the result of the high degree of similarity between the sequences aligned by the GAP program - 52% for AVALD and 56% for MVARG. The gaps introduced in these alignments may correspond to the differences noted between the secondary structural elements predicted by PREDICT. If higher level structure is a better measure of homology between proteins than simple sequence alignments then the archaebacterial enzyme would seem to be closely homologous in functionally important regions to its eukaryotic counterpart isolated from yeast. These proposed secondary structures

also fit in well with the predictions for all the previously sequenced pyruvate kinases based on the known three-dimensional structure of the cat muscle enzyme (Fothergill-Gilmore and Michels, 1992). The peptide AVALD falls into the alignment close to the start of a large sheet region between amino acids 96 and 137 (cat muscle enzyme numbering) at the boundary between domains A and B of the enzyme. The peptide MVARG falls into a region of alternating helix-sheet structure between residues 290 and 403 approaching the boundary between domain B and the second part of domain A. As the active site of the enzyme is known to be located between domains A and B, it seems likely that the structure of the pocket in that region should be very well conserved. It appears from this study that the archaebacterial enzyme does indeed maintain the structural pattern characteristic of the site.

The other, non-conserved peptides were also subjected to secondary structure predictions. The pattern for the peptide SVRTA was sheet-turn-helix-sheet whilst that for the proposed N-terminal peptide KLARC was sheet-turn-sheet-turn-sheet-turn-sheet-turn-sheet-helix. Since the alignments of both of these peptides to other pyruvate kinases are unsatisfactory, little information can be derived from their secondary structures.

Chapter 8: General Discussion

CHAPTER 8: GENERAL DISCUSSION

8.1 Growth of *T. acidophilum*

One of the first aims of the project was to set up a system for the culturing of archaebacterial (particularly thermophiles) in Edinburgh. This was successfully achieved, although the volume of the cultures produced was very small (2 litres). The identity of the cells cultured was verified using morphological and enzymatic methods. The latter study revealed that *T. acidophilum*, in common with other archaebacterial species, contains an isocitrate dehydrogenase which is specific for both NAD^+ and $NADP^+$ cofactors (Potter, 1993). This observation lends support to the hypothesis that the archaebacteria deserve separate taxonomic classification from both eukaryotes (which have two isocitrate dehydrogenases, one specific for each cofactor) and from eubacteria (which have, in most cases, only an $NADP^+$-linked enzyme).

8.2 Purification and characterisation of the pyruvate kinase

In order to evaluate pyruvate kinase as a phylogenetic marker for molecular evolutionary studies, the enzyme was purified from *T. acidophilum*. The protocol used for the purification was adapted from previous protocols used to purify pyruvate kinases from other sources and, after five steps, yielded microgram quantities of homogeneous, active enzyme from milligram quantities of cells. The pure enzyme was then physically and kinetically characterised and was found to closely resemble its mesophilic counterparts. At 60^oC, the archaebacterial pyruvate kinase was found to be functionally equivalent to eubacterial and eukaryotic enzymes whilst being significantly more stable at higher temperatures. The enzyme was found to be allosterically regulated, suggesting that it shares some structural properties (such as effector sites) with counterparts which are also regulated in like fashion, and evidence was also gathered which suggested that the enzyme may undergo thermal activation at around 50^oC.

The purified enzyme was also partially protein sequenced in order to allow the construction of oligonucleotide probes which would be

used for obtaining the gene for the protein. The N-terminus of the enzyme proved to be chemically blocked (a feature common amongst pyruvate kinases), necessitating its proteolytic digestion prior to internal peptide sequencing. Three peptide sequences were obtained by automated Edman degradation sequencing, one of which clearly corresponded to a conserved region of other pyruvate kinases identified by computer alignment studies. This suggested that the functional similarity to mesophilic enzymes displayed by the archaebacterial enzyme might be reflected in a close primary structural relationship.

8.3 Cloning strategy

The cloning of archaebacterial protein-encoding genes, using a variety of different approaches, is now well-documented in the literature. The construction of "mini-libraries", the use of oligonucleotides as probes, and a directional cloning approach were all important aspects of the strategy employed to clone the *T. acidophilum* pyruvate kinase gene. Oligonucleotides based on sequence information from two of the internal peptides derived as described above were used to identify the pyruvate kinase gene. This method was chosen as it did not require functional expression of the enzyme. Fragments of genomic DNA identified by the probes were then cloned and partially sequenced.

8.4 Sequence information

The derived partial amino acid sequences of the archaebacterial pyruvate kinase were compared by sequence alignment analyses with both eubacterial and eukaryotic enzyme sequences. The use of one program, GAP, gave alignments which suggested significant homology (around 30%) between the archaebacterial partial sequences corresponding to conserved regions of the enzyme and the complete mesophilic enzyme sequences. The multiple alignment program CLUSTAL, however, suggested that the homology score should be much lower - around 16%. The most reasonable interpretation of the data would probably be somewhere in between these extremes, with the archaebacterial protein retaining residues of structural and/or catalytic importance whilst lacking general sequence similarity to its eukaryotic and eubacterial counterparts. This combination of low identity scores with conservation of important residues has been reported for

comparisons between other archaebacterial and non-archaebacterial proteins [for examples, see Hensel *et al.*, 1989 and Sanangelantoni *et al.*, 1990].

An analysis of amino acid exchanges occurring between *T. acidophilum* and *S. cerevisiae* pyruvate kinase sequence stretches was carried out in order to identify possible thermophilic features of the archaebacterial protein. Since the archaebacterial sequence is not complete, however, no conclusions can be safely drawn although it is interesting to note that some predictions about changes to enhance the thermostability of a protein are fulfilled - increased numbers of proline residues, for example.

The sequence stretches obtained were also analysed in order to predict their secondary structure. Although the PREDICT program used is known to be unreliable (only 65% accurate), some interesting information was derived, particularly from the predictions of the peptides corresponding to conserved regions of the enzyme. It appears that the secondary structure corresponding to the active site region of pyruvate kinase, as determined from the known cat muscle enzyme three-dimensional structure, is essentially conserved in the archaebacterial enzyme. This lends weight to the concept that protein homologies are better detected by comparisons of higher level structure conservation than by simple sequence alignments.

8.5 Evolution of pyruvate kinase and its application as a phylogenetic marker

Fig. 1.10 shows the evolution rate for pyruvate kinase, based on the differences of the sequences of a limited set of organisms, for which the time at which they diverged was estimated as described in Fothergill-Gilmore and Michels (1992). The divergence times used were: primate - rodent/ungulate, 80 million years; mammal - bird, 290 million years; mammal - fish, 400 million years; vertebrate - invertebrate, 600 million years; animal - fungus - plant, 1000 million years; and eukaryote - prokaryote, 1800 million years. Whilst the majority of glycolytic enzymes seem to evolve at a fairly constant rate (following Wilson's molecular clock, 1985), the plot for pyruvate kinase is distinctly bi-phasic. This could be explained in a number of ways. For instance, some

sequences used in the analysis may not be representative of a phylogenetic group because of specific metabolic constraints within the organism from which the enzyme was isolated. Alternatively, anomalies can be explained by the occurrance of horizontal gene transfers between organisms, by the acquisition by the enzyme of other functions or even by the fact that the rate of evolution is not necessarily the same in every organism. Assuming that the curve in Fig. 1.10 is correct, however, the low sequence identity between the archaebacterial enzyme and those from eubacterial and eukaryotic sources would suggest that the enzyme should be very old indeed, possibly presaging the eukaryote - prokaryote division. This would lend support to the idea that archaebacteria are 'living fossils' and amongst the oldest of living organisms.

Given the low sequence identity between the pyruvate kinase of the archaebacterium and its counterparts from other kingdoms, it would seem inappropriate to utilise this enzyme as a phylogenetic marker. The identity scores (from multiple alignment analysis) are simply too low to produce sensible trees and would instead lead to the construction of a highly anomalous tree such as that constructed for glyceraldehyde-3-phosphate dehydrogenases (GAPDHs - Hensel *et al.*, 1989). There are three possible evolutionary histories for the archaebacterial enzyme. 1) the pyruvate kinase of archaebacteria, like their GAPDHs but unlike their phosphoglycerate kinases (Fabry *et al.*, 1990), has diverged from those of eubacteria and eukaryotes many millions of years before the divergence of the eubacterial and eukaryotic enzymes whilst retaining essential structural features. 2) an ancient lateral gene transfer event occurred between representatives of the eubacterial and eukaryotic kingdoms millenia after the original divergence of their ancestral pyruvate kinase from the archaebacterial enzyme. 3) the archaebacterial enzyme is not evolutionarily related to the enzymes found in the other two kingdoms at all, but is in fact a different protein which has, by a process of convergence, evolved to catalyse the same reaction. The fact that the archaebacterial enzyme possesses at least some regions which are recognisably homologous with other pyruvate kinases suggests that options 1 and 2 are more likely to be correct than is option 3.

8.6 Future work

Future work on this project will concentrate upon:

(a) Further characterisation of the *T. acidophilum* pyruvate kinase, particularly CD, NMR and fluorescence studies to ascertain whether the temperature dependent conformational change postulated here is real.

(b) Enhancement of the culturing facility for the cells with regard to the volume of the batches produced.

(c) Complete sequencing of the enzyme, alignment of the full sequence and complete secondary structural analysis.

(d) Expression of the cloned gene in *E. coli* which will allow the production of large quantities of the enzyme for three-dimensional structural analysis

(e) Purification and cloning of pyruvate kinases from other representative archaebacterial species.

Appendix 1: Purification and Partial Characterisation of a Protein of Unknown Function from *T. acidophilum*

APPENDIX 1: PURIFICATION AND PARTIAL CHARACTERISATION OF A PROTEIN OF UNKNOWN FUNCTION FROM *T. ACIDOPHILUM.*

A.1.1. Introduction

This chapter describes the isolation, using part of the purification protocol developed for the *T. acidophilum* pyruvate kinase, of a protein of unknown function from that archaebacterium. It also describes the physical characterisation of the protein and initial attempts to identify it using structural database searches.

A.1.2 Results

A.1.2.1 Purification of the protein of unknown function

During the development of the purification protocol for the pyruvate kinase of *T. acidophilum* (2.2.4), it was noted that several fractions of the eluent from the size-exclusion chromatography column step apparently contained only one protein. This was represented by a single band on an SDS-PAGE gel stained with Coomassie blue of relative molecular mass 20k. Such a gel, showing several fractions eluted from the Superose 6 column, is shown in Fig. A.1.1. From a gel of this type, densitometry analysis suggested that the protein amounted for approximately 3-4% of total cell protein. 5 mg in total of the protein was purified in this fashion.

The purity of the protein preparation was confirmed by H.P.L.C. on an Applied Biosystems 130A microbore separation system. The column used was an Aquapore RP-300 (7 mm particle size; 2.1 mm x 30 mm) and it was eluted with a linear 8-80% (v/v) gradient of acetonitrile in aqueous 0.1% TFA. Elution of the sample from the column was followed at 220 nm. The elution profile from the H.P.L.C. column is shown in Fig. A.1.2.

A.1.2 Characterisation of the protein of unknown function

The protein, designated Ta20kp, was physically characterised using amino acid analysis, amino acid sequencing and mass spectrometry techniques. Both amino acid analysis and amino acid sequencing were

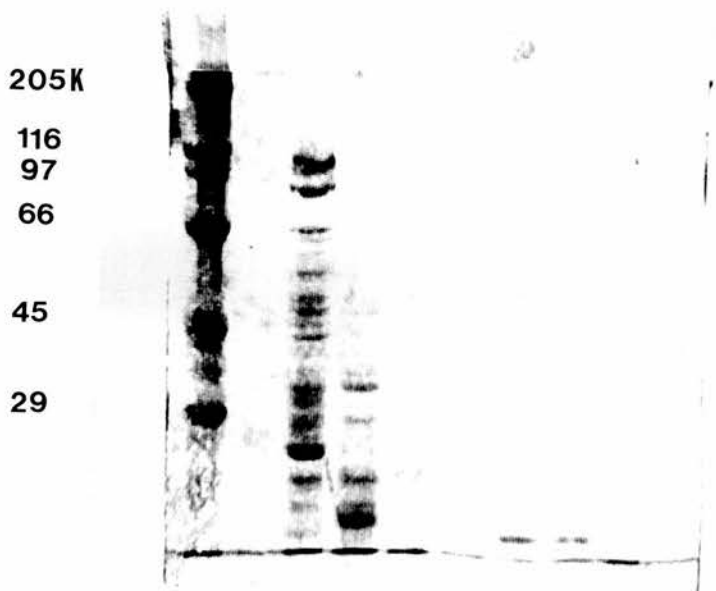


Figure A1.1: SDS-polyacrylamide gel electrophoresis of purified Ta20Kp. Lanes (L-R): High molecular weight markers (Sigma); *T. acidophilum* Superose 6 protein fraction; *T. acidophilum* Superose 6 protein fraction; pure Ta20Kp; pure Ta20Kp.

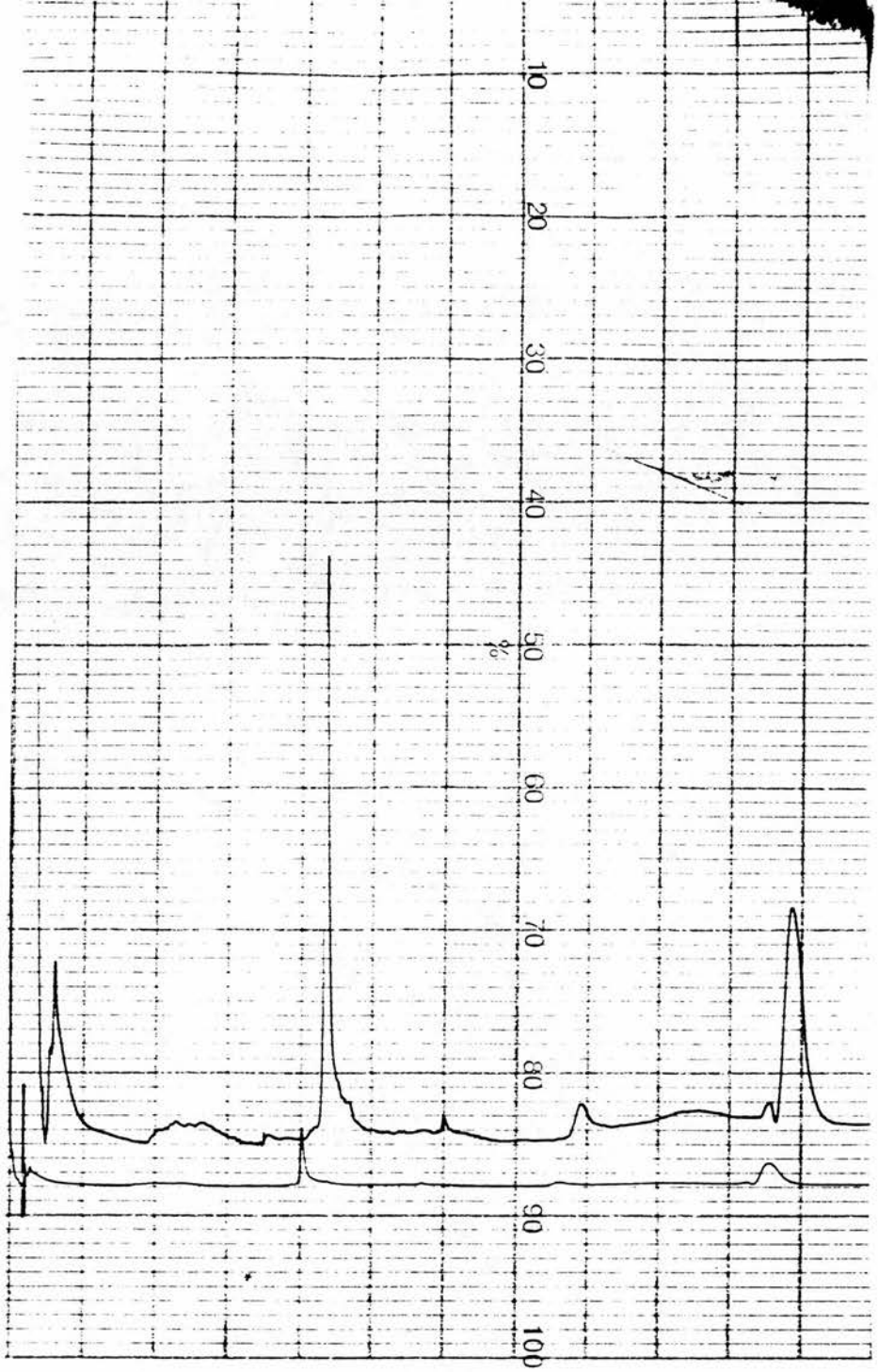


Figure A.1.2: Elution profile of purified Ta20Kp from the Aquapore RP-300 h.p.l.c. column.

performed on the Ta20kp sample in a manner equivalent to that used on the pyruvate kinase of *T. acidophilum* (2.2.12 - 2.2.14). The only variation was that Ta20kp was found to possess an N-terminus free of blocking groups and hence the sequence determined for Ta20kp is N-terminal. Fig. A.1.3. shows the output from the amino acid analyser. Table A.1.1 shows the derived amino acid composition of Ta20kp. Fig. A.1.4 shows the first 20 amino acids of the N-terminus of the protein.

Mass spectrometry of the protein was performed using a Lasermat Mass Analyser (Finnigan Mat) according to the method described in 2.2.31. This analyser is a matrix-assisted laser desorption/ionisation 'time of flight' device. Approximately 10 pmol of the Ta20kp were loaded onto the matrix substance (a 1000 fold molar excess) in the sample holder. The mixture was allowed to dry and form a crystalline deposit which was then irradiated with a short pulse of UV light from a laser. The matrix acts as a strong UV absorber and the energy from the laser causes the mixture of sample and matrix to volatilise, taking both into the gas phase. It is thought that ionised matrix molecules then transfer a proton to the sample molecules. The ions so formed are accelerated by a strong electrical field along the flight tube (maintained at a vacuum) of the instrument towards an electron multiplier detector. All the ions have the same energy and hence their velocities through the flight tube are related to their charge/ mass ratios. Higher mass ions travel slower than smaller ones. The protein-related ions produced are predominantly charge states z=+1 $(MH)^{+}$ and z=+2 $(MH2)^{++}$ (where z = charge and M is the sample) so that the spectra produced are relatively simple to interpret.

A typical spectrum is shown in Fig. A.1.5. This spectrum represents a summation of 11 laser shots at an accelerating voltage of 20018 V. As can be seen from the spectrum, the relative molecular mass of the protein has been more accurately determined to be approximately 17.7k.

A.1.3 Identification of the protein of unknown function

The 20 N-terminal amino acids determined for the protein were fed into the NBRF and SWISSPROT databases using the FASTA alignment program (Daresbury Vax system). This program produces similar alignments to those produced by the CLUSTAL package described in chapter 1. The system of penalties is also analogous. The alignment

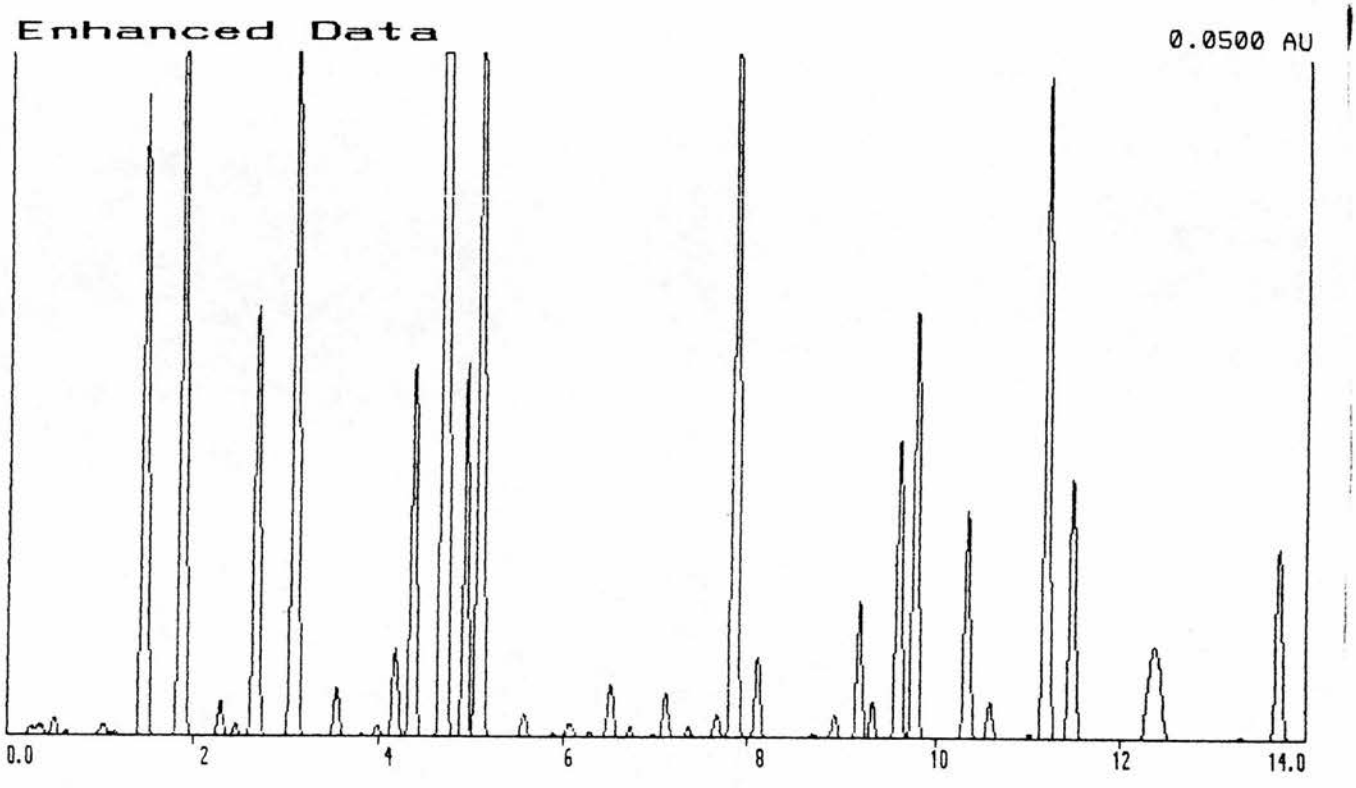


Figure A.1.3: Amino acid analysis of the Ta20Kp. The chromatogram shown was produced and interpreted as described for the *T. acidophilum* pyruvate kinase (see chapter 5). The composition derived from the analysis is tabulated in Table A.1.1.

amino acid	peak area	no. of residues/molecule
Asx	779148	20
Glx	920582	24
Ser	551256	12
Gly	906570	21
His	59370	1.5
Arg	107866	2
Thr	455134	10
Ala	2146228	36
Pro	378714	7
Tyr	51248	1
Val	896866	15.5
Met	93058	1.5
Ile	349330	18
Leu	503726	9
Phe	44928	1
Lys	808468	9

Table A.1.1: Amino acid composition of Ta20Kp. Peak areas are shown in uAU. Actual amino acid numbers were calculated on an estimated relative molecular mass of 20K.

	Cycle									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Ala	**243**	9	1	1	0	0	1	**196**	0	0
Arg	3	1	0	0	0	0	4	2	0	1
Asn	2	2	0	0	0	0	12	4	0	1
Asp	0	12	2	0	0	**91**	8	2	0	**105**
Cys	1	0	0	0	0	0	0	0	0	1
Glu	3	6	0	0	0	4	0	1	0	2
Gln	0	0	0	0	0	0	1	1	0	0
Gly	5	0	0	0	**176**	0	10	2	6	0
His	0	1	0	0	0	0	0	0	0	0
Ile	5	4	2	0	0	0	0	0	11	2
Leu	1	0	0	10	6	2	8	4	1	2
Lys	1	0	**136**	5	0	0	**147**	0	4	10
Met	35	2	0	0	0	0	1	2	1	1
Phe	0	0	0	1	1	0	2	12	0	1
Pro	20	1	0	0	12	2	1	0	**164**	12
Ser	10	0	0	0	0	0	1	0	0	1
Thr	11	1	0	0	8	2	1	1	0	1
Trp	8	0	3	1	2	0	11	0	0	3
Tyr	3	1	1	0	0	0	0	1	0	0
Val	1	**250**	14	**196**	0	0	0	15	3	4

Figure A1.4: Sequence report for Ta20Kp. The sequence is that of the N-terminal 20 amino acids. The values shown are lag corrected and in pmol. The deduced sequence and that of the corresponding oligo nucleotide probe are also shown.

	Cycle									
	11	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
Ala	7	0	**123**	0	0	1	0	0	0	5
Arg	0	2	0	0	0	0	0	0	0	0
Asn	3	0	1	0	0	0	**69**	0	0	0
Asp	0	0	0	0	**57**	1	4	0	0	0
Cys	0	0	0	0	0	0	0	0	0	0
Glu	0	**56**	2	0	0	0	1	0	0	0
Gln	0	0	3	0	0	1	0	0	0	0
Gly	4	0	4	0	0	0	0	3	3	0
His	0	0	0	0	1	0	0	0	0	0
Ile	10	0	1	1	0	0	0	0	0	1
Leu	1	0	0	0	0	5	0	**85**	4	0
Lys	0	0	0	0	0	2	0	0	**18**	0
Met	1	0	0	0	9	1	0	0	0	**40**
Phe	**130**	0	0	0	1	2	0	0	0	0
Pro	0	0	0	**95**	27	0	4	0	0	0
Ser	0	0	0	1	0	0	0	0	0	1
Thr	0	0	0	0	0	**65**	1	0	2	0
Trp	1	0	0	1	2	1	0	0	9	1
Tyr	0	4	1	0	7	0	0	0	0	0
Val	0	0	0	0	3	6	0	0	0	0

Deduced amino acid sequence: **AVKVGDKAPDFEAPDTNLKM**

Corresponding probe sequence: 5'-GCAGTGAAGGTGGGCGATAAGGCACCAGAT
TTCGAGGCACCAGATACGAACCTCAAGATG-3'

produced 40 possible matches, 11 of which were N-terminal. The full list of possibilities is shown in Table A.1.2.

A.1.4 Preliminary cloning experiments

In an attempt to improve this situation by providing the entire sequence of the gene for alignment analysis, cloning experiments were started using the N-terminal amino acids and the citrate synthase codon usage table (Sutherland, 1991) as a guide for constructing an oligonucleotide probe (O3). The sequence of the probe is shown in Fig. A.1.4.

Genomic DNA was isolated from *T. acidophilum* using a method based on that of Bowen et al. (1988, section 2.2.19) and 25 ng was restricted using the enzymes EcoRI and HinDIII. The digest was separated on an agarose gel which was then Southern blotted onto Hybond N membrane (2.2.22). The membrane was subsequently hybridised to fluorescein-labelled O3 probe and detected using the ECL system (Amersham, 2.2.23). A DNA fragment of size 3kb was detected by the probe at a stringency temperature of 45^{o}C - see Fig. A.1.6.

A.1.3 Discussion

From the results obtained to date, it is impossible to say what the protein is, especially given the propensity for archaebacterial proteins to not resemble their mesophilic counterparts in primary structure. 40 possibilities have been identified, however, some of which (sperm whale haemoglobin alpha-chain, for example) can be easily discounted. Obviously, a positive identification will be facilitated by the use of the entire sequence in the alignment program and hence the future work on this project should be directed towards continuing the cloning work started here. The cloning strategy used could closely shadow that used for the pyruvate kinase (chapter 6) i.e. use of directional cloning to amplify and sequence the 3kb EcoRI/HinDIII fragment identified by the N-terminus specific probe O3. The protein would seem to be an important one for *T. acidophilum* as it comprises up to 4% of total cell protein and hence identifying it may shed some further light onto the biochemistry of this unusual organism.

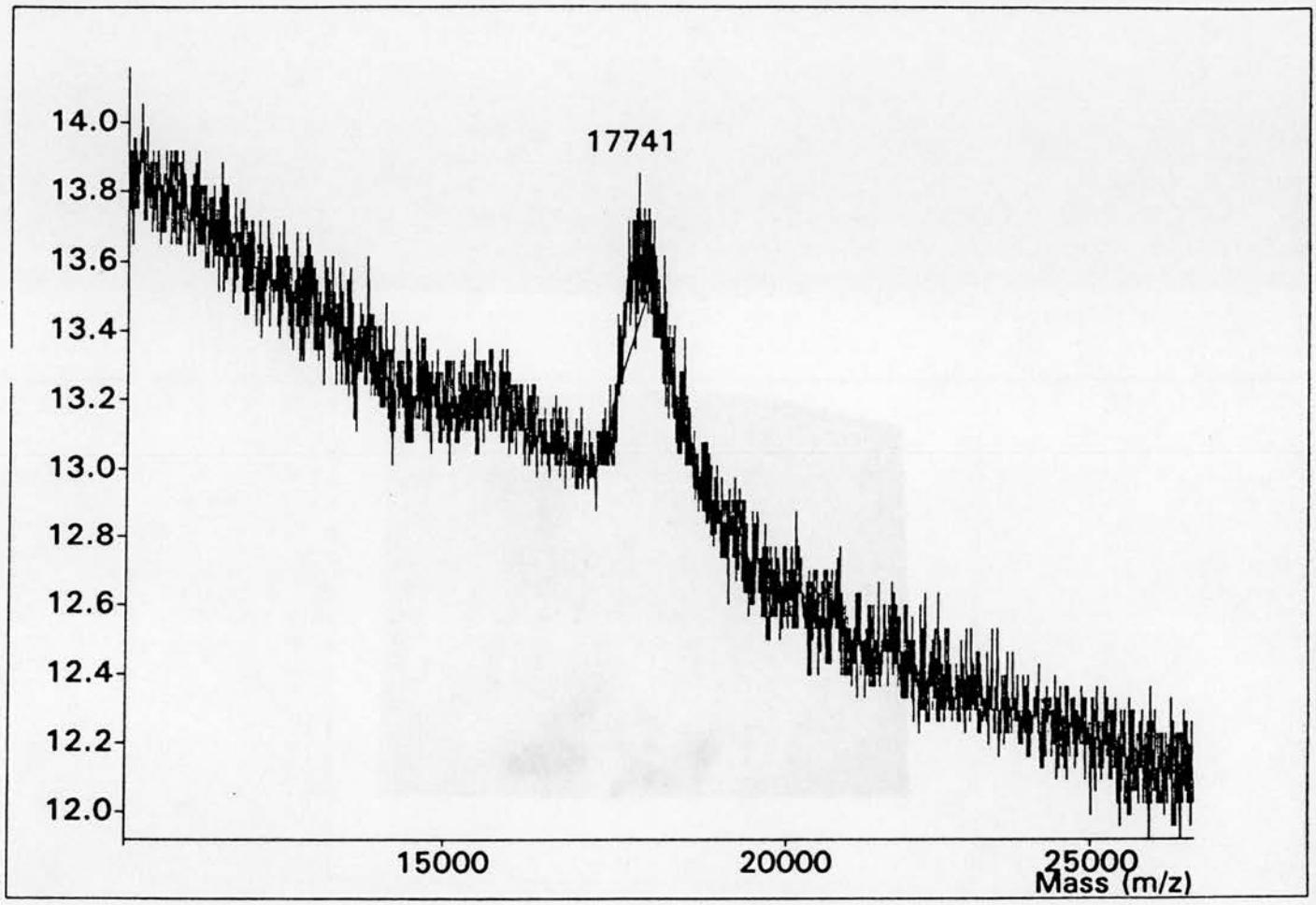


Figure A.1.5: Laser desorption mass spectrometry of the Ta20Kp. The Lasermat mass analyser was previously calibrated using cytochrome C. The relative molecular mass of the protein is indicated above the peak.

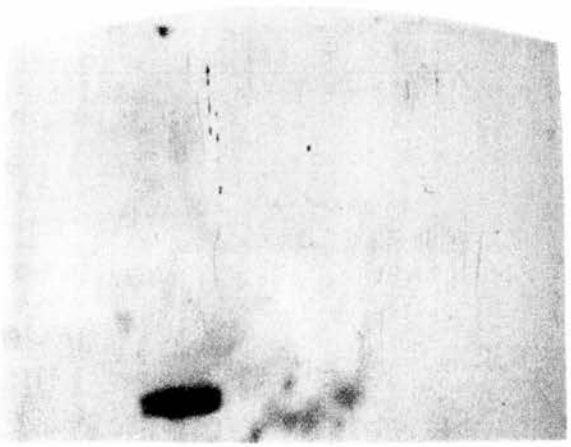


Figure A1.6: Hybridisation of fluorescein-labelled O3 to a Southern blot of an EcoRI/HinDIII double digest of *T. acidophilum* DNA. The blot was developed using the Amersham ECL system. The band size is 3 kb.

	protein
1	M protein (*Streptococcus*)
2	Vitellogenin (frog)
3	1-phosphatidyl-inositol bisphosphate phosphatase
4	Subtilisin (*Bacillus*)
5	Glucose repressible alcohol dehydrogenase
6	Major DNA-binding protein (Herpes simplex)
7	Ig heavy chain precursor (Mouse)
8	Phosphoribosylglycinamide formyltransferase
9	Hypothetical protein
10	Hypothetical protein L
11	DNA-binding protein (Herpes simplex)
12	Topoisomerase 1 (yeast)
13	Twitchin (*C. elegans*)
14	DNA-binding protein (Herpes simplex)
15	Ig heavy chain precursor W3129
16	Ig light chain V region (anticyclosporine)
17	Ig heavy chain V region (anticyclosporine)
18	Gene UL29 protein
19	phnP protein plasmid BW120 (*E. coli*)
20	Subtilisin BPN' precursor (*Bacillus*)
21	Ig heavy chain V region (H158-89H4 -mouse)
22	Ig heavy chain V region (H37-62-mouse)
23	Threonine dehydratase
24	Ig heavy chain V region (914-mouse)
25	ilvA leader peptide (*E.coli*)
26	Ig heavy chain V region (anticyclosporine)
27	Threonine dehydratase
28	Haemoglobin alpha chain (Sperm whale)
29	Dolichyldiphosphooligosaccharide protein g
30	Ig heavy chain V region (anticyclosporine)
31	Ig heavy chain V region (H35-C6-mouse)
32	Arginosuccinate synthase
33	Nitrogenase
34	Yellow protein (fruit fly)
35	M polyprotein precursor (*Nephropathia*)
36	IgG Fc receptor precursor type III-1
37	Protein disulphide isomerase
38	Glutathione reductase (NADPH)
39	UDP-glucose hydrolase precursor
40	Hypothetical UL105 protein (human cytomegalovirus)

AC. no.	identical aa	conserved aa	window	% ident.
Pir2:A28549	7	7	18	38.9
Pir2:S03124	6	9	16	37.5
Pir2:A28807	7	5	12	58.3
Pir1:Subsd	8	9	19	42.1
Pir2:A24534	6	3	9	66.7
Pir1:Dnbev1	6	13	20	30.0
Pir1:Hvms84	5	11	19	26.3
Pir2:I29326	8	8	19	42.1
Pir2:S02053	7	5	15	46.7
Pir2:A28667	8	6	18	44.4
Pir1:Dnbehf	6	13	20	30.0
Pir2:A23161	5	4	10	50.0
Pir2:S07571	6	8	15	40.0
Pir1:Dnbeks	6	13	20	30.0
Pir2:A25912	6	10	19	31.6
Pir3:Ph0095	5	11	19	26.3
Pir3:Ph0094	5	11	19	26.3
Pir3:B30085	6	13	20	30.0
Pir2:H35719	6	5	14	42.9
Pir1:Subsn	5	12	19	26.3
Pir2:F27888	5	10	19	26.3
Pir2:C27888	5	10	19	26.3
Pir1:Dwects	8	8	18	44.4
Pir1:Hvms91	5	10	19	26.3
Pir3:E26570	8	8	18	44.4
Pir3:Ph0097	5	10	19	26.3
Pir3:Jt0278	8	7	18	44.4
Pir2:A25728	7	4	11	63.6
Pir2:A30007	5	6	12	41.7
Pir3:Ph0096	5	10	19	26.3
Pir2:E27888	5	10	19	26.3
Pir2:C28180	7	9	19	36.8
Pir1:Nibcat	6	7	16	37.5
Pir2:A25696	6	3	10	60.0
Pir1:Gnvune	8	6	16	50.0
Pir2:Ju0284	5	8	14	35.7
Pir1:Isrtss	4	7	12	33.3
Pir1:Rdhuu	6	5	12	50.0
Pir1:Yxecug	5	6	12	41.7
Pir1:Qqbek2	7	10	20	35.0

Table A.1.2: Database search output using FASTA alignment program. The 20 amino acid sequence stretch of Ta20Kp was used for the search. % identity values are based on the window size which is indicated in each case. The ACno. is the accession code for each protein in the Protein Information Resource (Pir) database.

References

References

Ambler, R.P. (1985) Protein sequencing and taxonomy. pp 307-335 in Goodfellow, M., Jones, D. & Priest, F.G. (eds.) Computer assisted bacterial systematics, Academic Press, London.

Amils, R., Ramirez, L., Sanz, J.L., Marin, I., Pisabarro, A.G. & Urena, D. (1989) Can. J. Microbiol. 35, 141-147

Andreesen, J.R. & Gottschalk, G. (1969) Arch. Microbiol. 69, 160-170

Argos, P., Rossman, M.G., Grau, U.M., Zuber, H., Frank, G. & Tratschin, J.D. (1979) Biochemistry 18, 5698-5703

Arnone, M.I., Birolo, L., Giamberini, M., Cubellis, M.V., Nitti, G., Sannia, G. & Marino, G. (1992) Eur. J. Biochem. 204, 1183-1189

Blair, J.B. & Harman, D.L. (1986) Fed. Proc. 45, 1657

Bolton, E.T. & McCarthy, B.J. (1962) Proc. Natl. Acad. Sci. USA 48, 1390-1397

Bowen, D., Littlechild, J.A., Fothergill, J.E., Watson, H.C. & Hall, L. (1988) Biochem. J. 254, 509-517

Bradford, M.M. (1976) Anal. Biochem. 72, 248-254

Brock, T.D. (1985) Science 230, 132-138

Bright, J.R., Byrom. D., Danson, M.J., Hough, D.W. & Towner, P. (1993) Eur. J. Biochem. 205, 549-554

Brown, J.W., Daniels, C.J. & Reeve, J.N. (1986) CRC Crit. Rev. Microbiol. 16, 287-338

Bucher, T. & Pfleiderer, G. (1955) Meth. Enzymol. 1, 435-440

Budgen, N. & Danson, M.J. (1986) FEBS Lett 196, 207-210

Burgess, A.W., Ponnuswamy, P.K. & Scheraga, H.A. (1974) Isr. J. Chem. 12, 239-286

Burke, R.L., Tekamp-Olson, P. & Najarian, R. (1983) J. Biol. Chem. 258, 2193-2201

Chatton, E. (1938) Titres et travaux scientifique (1906-1937) de Edouard Chatton, Sottano, Sete, 1938

Chothia, C. (1988) Nature 333, 598-599

Chou, P.Y. & Fasman, G.D. (1974) Biochemistry 13, 222-245

Christiansen, C., Freundt, E.A. & Black, F.T. (1975) Int. J. Syst. Bacteriol. 25, 99-101

Collins, R.A., McNally, T. & Fothergill-Gilmore, L.A. (1992) Biochem. Soc. Trans. 21, 63S

Cooper, R.A. (1986) *in* Carbohydrate Metabolism in Cultured Cells. *edited by* M.J. Morgen. Plenum Press, New York and London. pp.461-491

Cornish-Bowden, A. (1977) J. Theor. Biol. 65, 735-742

Cornish-Bowden, A. (1979) Fundamentals of enzyme kinetics, Butterworth & Co., London, pp 152-154

Cowan, D.A. (1992) pp 1-43 in Danson, M.J., Hough, D.W. & Lunt, G.G. (eds.) The archaebacteria: biochemistry and biotechnology, Portland press, London.

Creighton, T.E. (1984) Proteins: structures and molecular properties, W.H. Freeman, N.Y.

Danson, M.J. (1988) Adv. Microbiol. Physiol. 29, 165-231

Danson, M.J. & Wood, P.A. (1984) FEBS Lett. 172, 289-293

Darland, G., Brock, T.D., Samsonoff, W. & Conti, S.F. (1970) Science 170, 1416-1418

Davies, J.M., Rea, P.A. & Sanders, D. (1991) FEBS Lett. 278, 66-88

Dayhoff, M.O., Schwartz, R.M. & Orcutt, B.C. (1978) pp 345-358 in Atlas of protein structure vol.5 suppl.3, National biomedical research foundation, Washington D.C.

de Graaf, L. & Visser, J. (1988) Curr. Genet. 14, 553-560

DeLange, R.J., Green, G.R. & Searcy, D.G. (1981a) J. Biol. Chem. 256, 900-904

DeLange, R.J., Green, G.R. & Searcy, D.G. (1981b) J. Biol. Chem. 256, 905-911

De Rosa, M., Gambacorta, A., Nicolaus, B., Giardina, P., Poerio, E. & Buonocore, V. (1984) Biochem. J. 224, 407-414

Dufton, M.J. & Hider, R.C. (1977) J. Mol. Biol. 115, 177-193

Duncan, D., McGowan, J. & Price, N.C. (1989) Biochem. Soc. Trans. 17, 560-561

Edman, P. & Begg, G. (1967) Eur. J. Biochem. 1, 80-91

Eigen, M. (1967) Nobel Symposium 5, 333

Eliopoulos, E.E., Geddes, A.J., Brett, M., Papin, D.J.C. & Findlay, J.B.C. (1982) Int. J. Biol. Macromol. 4, 263-268

Fabry, S. & Hensel, R. (1987) Eur. J. Biochem. 165, 147-155

Fabry, S., Heppner, P., Dietmaier, W. & Hensel, R. (1990) Gene 91, 19-25

Facchiano, F., Ragone, R., Porcelli, M., Cacciapuoti, G. & Colonna, G (1992) Eur. J. Biochem. 204, 473-482

Farber, G.K. & Petsko, G.A. (1990) Trends in Biochem. Sci. 15, 228-234

Fewson, C.A. (1986) Biochem. Educ. 14 (3), 103-115

Fontana, A. (1988) Biophys. Chem. 29, 181-193

Forterre, P. (1992) in Frontiers of Life, Editions Frontieres, Gif sur Yvette, France (in press)

Fothergill-Gilmore, L.A. & Michels, P.A.M. (1992) Prog. Biophys. Molec. Biol. 59, 105-235

Franks, F. (1991) in Meeting book, Protein stabilisation course, Cambridge 1991

Fuhrman, J.A., McCallum, K. & Davis, A.A. (1992) Nature 356, 148-149

Garnier, J., Osguthorpe, D.J. & Robson, B. (1978) J. Mol. Biol. 120, 97-120

Geisow, M.J. & Roberts, R.D.B. (1980) Int. J. Biol. Macromol. 2, 387-389

Gest, H. & Schopf, J.W. (1983) pp. 135-148 *in* Earth's Earliest Biosphere. *edited by* J.W. Schopf. Princeton University Press, Princeton.

Gross, E. & Witkop, B. (1962) J. Biol. Chem. 237, 1856-1860

Grossebuter, W., Hartl, T., Gorisch, H. & Stezowski, J.J. (1986) Biol. Chem. Hoppe-Seyler 367, 457-463

Grunstein, M. & Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965

Guderley, H., Fournier, P. & Auclair J-C. (1989) J. Theor. Biol. 140, 205-220

Hall, E.R. & Cottam, G.L. (1978) Int. J. Biochem. 9, 785-793

Hall, E.R. & Hobson, P.N. (1988) Anaerobic digestion 1988, Pergamon press, Oxford

Hampp, N., Brauchle, C. & Oesterhelt, D. (1990) Biophys. J. 58, 83-93

Hanish, J. & McClelland, M. (1988) Gene Anal. Tech. 5, 105

Hayes, J.D., Kerr, L.A. & Cronshaw, A.D. (1989) Biochem. J. 264, 437-445

Hennig, M., Schlesier, B., Dauter, Z., Pfeffer, S., Betzel, C., Hohne, W.E. & Wilson, K.S. (1992) FEBS Lett. 306, 80-84

Hensel, R., Zwickl, P., Fabry, S., Lang, J. & Palm, P. (1989) Can. J. Microbiol. 35, 81-85

Higgins, D.G. & Sharp, P.M. (1988) Gene 73, 237-244

Hill, R. (1925) Proc. R. Soc. B100, 419

Hough, D.W. & Danson, M.J. (1989) Lett. Appl. Microbiol. 9, 33-39

Hunger, H.D. & Reinbothe, H. (1974) J. Chromatogr. 97, 273-276

Hunsley, J.R. & Suelter, C.H. (1969) J. Biol. Chem. 244, 4815-4818

Jones, W.J., Nagle, D.P. & Whitman, W.B. (1987) Microbiol. Rev. 51, 136-177

Klibanov, A.M. (1989) Trends in Biochem. Sci. 14, 141-144

Klump, H., Di Ruggiero, J., Kessel, M., Park, J-B., Adams, M.W.W. & Robb, F.T. (1992) J. Biol. Chem. 267, 22681-22685

Koshland, D.E., Nementhy, G. & Filmer, D. (1966) Biochemistry 5, 365-385

Kuczenski, R. & Suelter, C.H. (1972) Biochemistry 9, 2043

Laemmli, U.K. (1970) Nature 227, 680-685

Lake, J.A. (1988) Nature 331, 184

Lake, J.A. (1989) Can. J. Microbiol. 35, 109-118

Lake, J.A. (1991) Trends in Biochem. Sci. 16, 289-290

Langworthy, T.A., Smith, P.F. & Mayberry, W.R. (1972) J. Bacteriol. 112, 1193-1200

Larsen, H. (1981) pp. 985-994 in The Prokaryotes (Starr, M.P., Stolp, H., Trouper, H.G., Balows, A. & Schlegel, H.G. eds.) vol.1, Springer Verlag, Berlin

Lenstra, J.A., Hofsteenge, J. & Beintema, J.J. (1977) J. Mol. Biol. 109, 185-193

Lim, V.I. (1974) J. Mol. Biol. 88, 873-894

Lonberg, N. & Gilbert, W. (1983) Proc. Natl. Acad. Sci. USA 80, 3661-3665

Lone, Y.C., Simon, M-P., Kahn, A. & Marie, J. (1985) FEBS Lett. 195, 97-100

Marchalonis, J.J. & Weltman, J.K. (1971) Comp. Biochem. Physiol. 38B, 609

Margolin, A.L., Tai, D.F. & Klibanov, A.M. (1987) J. Am. Chem. Soc. 109, 7885-7887

Margulis, L. & Guerrero, R. (1991) New Scientist , 46-50

Matsuhashi, M., Dietrich, C.P. & Strominger, J.L. (1965) Proc. Natl. Acad. Sci. USA 54, 587-592

Mayr, E. (1991) Nature 348, 491

McLachlan, A.D. (1977) Int. J. Quantum Chem. 12, Suppl. 1, 371-385

McNally, T. (1990) Ph.D. Thesis, University of Edinburgh

McNally, T. & Fothergill-Gilmore, L.A. (1990) Biochem. Soc. Trans. 19, 258

McNally, T., Purvis, I.J., Fothergill-Gilmore, L.A. & Brown, A.J.P. (1989) FEBS Lett. 247, 312-316

Mitchell, W.M. (1977) Meth. Enzymol. 47, 165-170

Monod, J., Wyman, J. & Changeaux, J.P. (1965) J. Mol. Biol. 12, 88-118

Muirhead, H. (1987) pp. 114-186 in Jurnak, F.A. & McPherson, A. (eds.) Biological macromolecules and assemblies vol.3

Muirhead, H., Clayden, D.A., Barford, D., Lorimar, C.G., Fothergill-Gilmore, L.A., Schlitz, E. & Schmitt, W. (1986) EMBO J. 5, 475-481

Murcott, T.H.L., McNally, T., Allen, S.C., Fothergill-Gilmore, L.A. & Muirhead, H. (1991) Eur. J. Biochem. 198, 513-519

Nagano, K. (1973) J. Mol. Biol. 75, 401-420

Neumann, H., Grierl, A., Tu, J., Leibrock, J., Staiger, D. & Zillig, W. (1983) Mol. Gen. Genet. 192, 66-72

Noguchi, T., Inoue, H. & Tanaka, T.(1986) J. Biol. Chem. 261, 13807-13812

Noguchi, T., Yamada, K., Inoue, H., Matsuda, T. & Tanaka, T. (1987) J. Biol. Chem. 262, 1-5

Pai, S.R. & Bird, R.C. (1991) Genetic analysis: Techniques and applications 8, 214-216

Pawluk, A., Scopes, R.K. & Griffiths-Smith, K. (1988) Biochem. J. 238, 275-281

Perutz, M.F. (1978) Science 201, 1187-1191

Perutz, M.F. & Raidt, H. (1975) Nature 255, 256-259

Pon, N.G. & Bondar, R.J.L. (1967) Anal. Biochem. 19, 272-279

Post, F.J. & Al-Harjan, F.A. (1988) Syst. Appl. Microbiol. 11, 97-101

Potter, S. (1992a) The Biochemist 14(1), 21-24

Potter, S. (1992b) The Biochemist 14(3), 16-20

Potter, S. (1992c) The Biochemist 14(5), 10-13

Potter, S. (1993) Can. J. Microbiol. 39, (in press)

Potter, S. & Fothergill-Gilmore, L.A. (1992a) Biochem. Soc. Trans. 20, 11S

Potter, S. & Fothergill-Gilmore, L.A. (1992b) FEMS Microbiol. Lett. 94, 235-240

Potter, S. & Fothergill-Gilmore, L.A. (1993) Biochem. Educ. 21 (in press)

Potter, S., Johnston, N. & Fothergill-Gilmore, L.A. (1993) Biochem. Soc. Trans. 21, 64S

Prangishvilli, D., Zillig, W., Gierl, A., Biesert, L. & Holz, I. (1982) Eur. J. Biochem. 122, 471-477

Ragone, R., Facchiano, F., Cacciapuoti, G., Porcelli, M. & Colonna, G. (1992) Eur. J. Biochem. 204, 483-490

Russell, A.J. & Klibanov, A.M. (1988) J. Biol. Chem. 263, 11624-11626

Sakai, H., Suzuki, K. & Imahori, K. (1986) J. Biochem. 99, 1157-1167

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1990) Molecular cloning. A laboratory manual, 2nd edition, Cold Spring Harbor laboratory press

Sanangelantoni, A.M., Barbarini, D., Di Pasquale, G., Cammarano, P & Tiboni, O. (1990) Mol. Gen. Genet. 221, 187-194

Schmitz, G.G., Walter, T., Seibl, R. & Kessler, C. (1990) Anal. Biochem. 192, 222-231

Searcy, D.G. (1986) System. Appl. Microbiol. 7, 198-201

Searcy, D.G. & Whatley, F.R. (1984) Syst. Appl. Microbiol. 5, 30-40

Segel, I.H. (1975) pp 113-118 in Enzyme kinetics: behaviour and analysis of rapid equilibrium and steady state systems, Wiley, N.Y.

Segerer, A., Langworthy, T.A. & Stetter, K.O. (1988) System. Appl. Microbiol. 10, 161-165

Shewan, J.M. (1971) J. Appl. Bacteriol. 34, 299-315

Sioud, M., Baldarchi, G., Forterre, P. & de Redondo, A.M. (1987) Eur. J. Biochem. 169, 231-235

Smith, L.D., Stevenson, K.J., Hough, D.W. & Danson, M.J. (1987) FEBS Lett. 225, 277-281

Sneath, P.H.A. (1989) System. Appl. Microbiol. 12, 15-31

Somero, G. & Hochachka, P. (1968) Biochem. J. 110, 395

Southern, E.M. (1975) J. Mol. Biol. 98, 503-517

Speranza, M.L., Valentini, G., Iadarola, P., Stoppini, M., Malcovalti, M. & Ferri, G. (1989) Biol. Chem. Hoppe-Seyler 370, 211-216

Strominger, J. (1955) Biochim. Biophys. Acta 16, 616-618

Suzuki, Y. (1989) Proc. Japn. Acad. Ser. B. Phys. Biol. Sci. 65, 146-148

Sutherland, K.J. (1991) Ph.D. Thesis, University of Bath

Sutherland, K.J., Danson, M.J., Hough, D.W. & Towner, P. (1991) FEBS Lett. 282, 132-134

Tani, K., Fuji, H., Nagata, S. & Miwa, S. (1988) Proc. Natl. Acad. Sci. USA 85, 1792-1795

Tomlinson, G.A., Koch, T.K. & Hochstein, L.I. (1974) Can. J. Microbiol. 20, 1085-1091

Tu, J. & Zillig, W. (1982) Nucleic Acids Res. 10, 7231-7245

Vihinen, M. (1987) Prot. Engineer. 1, 477-480

Wakabayashi, S., Fujimoto, N., Wada, K., Matsubara, H., Kerscher, L. & Oesterhelt, D. (1983) FEBS Lett. 162, 21-24

Walker, J.E., Wonnacott, A.J. & Harris, J.I. (1980) Eur. J. Biochem. 108, 581-586

Watson, H.C. & Littlechild, J.A. (1990) Biochem. Soc. Trans. 18, 187-190

Weitzman, P.D.J. (1981) Adv. Microbiol. Physiol. 22, 185-244

Whittaker, R.H. (1959) Q. Rev. Biol. 34, 210-226

Wilson, A.C. (1985) Scientific American 248, 148-157

Winnacker, E.L. (1987) "From genes to clones", VCH Verlagsgesellschaft (Weinheim, Germany)

Woese, C.R. (1981) Scientific American 244, 94-106

Woese, C.R. (1987) Microbiol. Rev. 51, 221-271

Woese, C.R. & Fox, G.E. (1977) Proc. Natl. Acad. Sci. USA 74, 5088-5090

Woese, C.R., Kandler, O. & Wheelis, M.L. (1990) Proc. Natl. Acad. Sci. USA 87, 4576-4579

Yun, S-L., Aust, A.E. & Suelter, C.H. (1976) J. Biol. Chem. 251, 124-128

Zillig, W., Klenk, H-P., Palm, P., Leffers, H., Puhler, G., Gropp, F. & Garrett, R.A. (1989) Endocytobiosis & Cell Res. 6, 1-25

Appendix 2: Publications

APPENDIX 2: PUBLICATIONS

1: Papers

[1] POTTER, S. & FOTHERGILL-GILMORE, L.A. (1992). Purification and properties of pyruvate kinase from *Thermoplasma acidophilum*. FEMS Microbiol. Lett. 94, 235-239.

[2] POTTER, S. (1993). Evidence for a dual-specificity isocitrate dehydrogenase in the euryarchaeotan *Thermoplasma acidophilum*. Can. J. Microbiol. 39, 262-264.

[3] POTTER, S. & FOTHERGILL-GILMORE, L.A. (1993). Molecular evolution: the origin of glycolysis. Biochem. Educ. 21, 45-48.

2: Magazine articles

[4] POTTER, S. (1992). Through semantics to systematics - the trouble with archaebacteria. The Biochemist 14(1), 21-24.

[5] POTTER, S. (1992). Outrageous! The archaebacteria - phylogeny and biochemistry. The Biochemist 14(3), 16-20.

[6] POTTER, S. (1992). The abyss gazes also: the lives and times of the early archaebacteria. The Biochemist 14(5), 10-13.

3: Abstracts

[7] POTTER, S. & FOTHERGILL-GILMORE, L.A. (1992). Biochem. Soc. Trans. 20, 11S.

[8] POTTER, S., JOHNSTON, N. & FOTHERGILL-GILMORE, L.A. (1993). Biochem. Soc. Trans. 21, 64S.

FEMS Microbiology Letters 94 (1992) 235–240

Published by Elsevier

FEMSLE 04955

Purification and properties of pyruvate kinase from *Thermoplasma acidophilum*

Simon Potter and Linda A. Fothergill-Gilmore

Department of Biochemistry, University of Edinburgh, Hugh Robson Building, Edinburgh, UK

Received 6 April 1992
Revision received 23 April 1992
Accepted 27 April 1992

Key words: Pyruvate kinase; Archaebacteria; Enzyme purification; Amino acid sequence; *Thermoplasma acidophilum*

1. SUMMARY

Thermoplasma acidophilum is a thermoacidophilic archaebacterium occupying a paradoxical place in phylogenetic trees (phenotypically it is a thermoacidophile but phylogenetically it classifies with the methanogens). To better understand its phylogeny, the pyruvate kinase from this organism is being investigated as a molecular marker. The enzyme has been purified and has a native M_r of 250000. It consists of four, apparently identical subunits each of M_r 60000. No remarkable kinetic differences have been found between this thermophilic enzyme and its mesophilic counterparts other than its greater thermostability. Its amino acid composition has been determined and some partial sequencing has been done.

Correspondence to: S. Potter, Department of Biochemistry, University of Edinburgh, Hugh Robson Building, George Square, Edinburgh EH8 9XD, UK.

2. INTRODUCTION

Pyruvate kinase (ATP-pyruvate 2-O-phosphotransferase, EC 2.7.1.40) catalyses the essentially irreversible transphosphorylation from phosphoenolpyruvate (PEP) to ADP, a reaction that requires magnesium and potassium ions. The enzyme has been extensively studied because of its importance in controlling the flux through glycolysis from fructose 1,6-bisphosphate down to pyruvate. Amino acid sequences [1] are known for pyruvate kinases isolated from several sources, and a crystal structure is available for the cat muscle enzyme [2]. The apparent ubiquity of the enzyme, together with the maintenance of function in all of the organisms from which it has been isolated, suggested its use as a molecular marker for the construction of phylogenetic trees. It would thereby be complementary to the RNA or ribosome phylogenies proposed by Woese [3] and Lake [4].

Phenotypically, with regards to glycolysis, the archaebacteria fall into two groups: (a) the

halophiles and methanogens, which possess an Entner-Doudoroff pathway and the enzymes in the lower part of the glycolytic pathway, and (b) the thermoacidophiles which utilise a non-phosphorylated pathway and possess only two of the common enzymes, enolase and pyruvate kinase [5]. These observations have led to the hypothesis that glycolysis evolved in the gluconeogenic direction (from the 'bottom up'). In order to investigate this hypothesis, the pyruvate kinase of the thermoacidophilic archaebacterium *Thermoplasma acidophilum* has been purified to homogeneity, characterised and partially sequenced. This study provides the first kinetic and sequence information about a glycolytic enzyme from a thermoacidophilic archaebacterium.

3. METHODS

3.1. Cell culture

Thermoplasma acidophilum was supplied as a freeze-dried culture by the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, FRG. Cells were cultured at 60°C and pH 2.0 (as recommended in the 1989 Catalogue of Strains provided by that organisation) in a temperature-controlled 2-l glass vessel.

3.2. Purification of the pyruvate kinase

The pyruvate kinase was purified from *Thermoplasma* crude extract by a series of chromatographic steps done at 45–50°C. Cells were collected by centrifugation and 200 mg were resuspended in 1.5 ml of lysis buffer (20 mM NaCl, 10 mM $MgCl_2$, 10 mM Tris · HCl (pH 7.5), containing 1,10-phenanthroline, E64C, 3,4-dichloroisocoumarin protease inhibitors; all 10 mM) and were homogenised with glass beads for 5 min at room temperature. The suspension was briefly centrifuged to remove cell debris, the crude extract was made up to 35% saturation with ammonium sulphate and left to stand at room temperature for 20 min with occasional mixing. The precipitate was removed by centrifugation for 2 min in a microcentrifuge and the supernatant was made up to 65% saturation with ammonium sulphate. After standing for 20 min, the mixture was centrifuged for 2 min and the supernatant was discarded. The pellet was resuspended in 1 ml of lysis buffer and this was dialysed overnight at 45°C against 1 l of lysis buffer (without the protease inhibitors). The dialysed sample was applied to a Mono-Q ion-exchange column (Pharmacia, 38 $mm^2 \times 42$ mm) and the column eluted with a linear gradient 0–300 mM NaCl in lysis buffer over 30 min at a flow rate of 0.5 ml/min. Fractions of 1 ml were collected, and those containing pyruvate kinase activity were pooled and applied to a Superose 6 gel filtration column (Pharmacia, 172 $mm^2 \times 285$ mm). The column was eluted isocratically with lysis buffer at a flow rate of 0.5 ml/min. Active fractions of 1 ml were pooled and applied to a column of 5′-AMP agarose (Sigma, 79 $mm^2 \times 50$ mm). The pure pyruvate kinase was eluted from the column with 10 mM phosphate buffer (pH 7.5) containing 10 mM ADP (Sigma).

3.3. Enzyme characterisation

The pyruvate kinase was characterised by enzyme assays, SDS-polyacrylamide gel electrophoresis, amino-acid analysis and partial sequencing. Two assays for the enzyme were used. For qualitative detection of pyruvate kinase activity during the purification procedure, the coupled assay of Bucher and Pfleiderer [6] was used at 45°C (a compromise temperature to allow high activity of the pyruvate kinase without complete extensive denaturation of the coupling lactate dehydrogenase). For kinetic studies however, this coupled assay could not be used and so the direct, spectrophotometric assay of Pon and Bondar [7] was employed at 60°C. For each assay the cuvette contained in 1 ml: $MgSO_4$, 7.2 mM; KCl, 7.2 mM; Tris, 0.05 M; pyruvate kinase, 1 μg and ADP and PEP in invariant concentrations. The pH of the buffer was fixed at 7.5.

Amino acid analysis was performed using an Applied Biosystems 420A Derivatizer with automated hydrolysis (Cronshaw, A.D., MacBeath, J.R.E., Shackleton, D.R., Fothergill-Gilmore, L.A. and Hulmes, D.J.S. unpublished results). PTC-amino acid derivatives were separated by reverse-phase HPLC and detected at 254 nm.

The N-terminus of the enzyme proved to be blocked and hence, prior to sequencing, the pyruvate kinase was proteolytically cleaved by clostripain (Sigma). The pyruvate kinase was reduced and carboxymethylated prior to digestion, and the protease was pretreated with 10 mM DTE for 2 h at 4°C. Digestion was done at pH 7.5 for 2 h at 37°C and at a protein/protease ratio of 125 : 1 and at 1.5 mg protein/ml of 20 mM ammonium bicarbonate. Resultant peptides were separated and purified using an Applied Biosystems 130A microbore separation system with an Aquapore RP-300 column (7 μm pore size; 2.1 mm $\times$ 30 mm). The purified peptides were then subjected to automated sequencing using an Applied Biosystems 477A instrument with a 120A on-line phenylthiohydantoin analyser [8].

4. RESULTS AND DISCUSSION

Pyruvate kinase purified by the above procedure was shown by SDS-PAGE to be homogeneous (Fig. 1). The increase in total activity noted during the procedure is possibly due to the presence of inhibitor molecules in the crude extract. The enzyme is a tetramer of subunits of M_r 60 000 (the subunit M_r was determined from SDS-PAGE gels and the tetramer M_r from the Superose 6 column). It has a specific activity of approximately 200 U/mg which is perhaps a little low for a pyruvate kinase (Table 2). This is not due to the assay conditions used (as the enzyme was incubated for 10 min at 60°C before the assay was started) but may be explained by partial cold denaturation of the enzyme — activity is negligible below 40°C. The inactivation is reversible if the enzyme is incubated at room temperature but freezing denatures it irreversibly. A purification table for the enzyme is shown in Table 1.

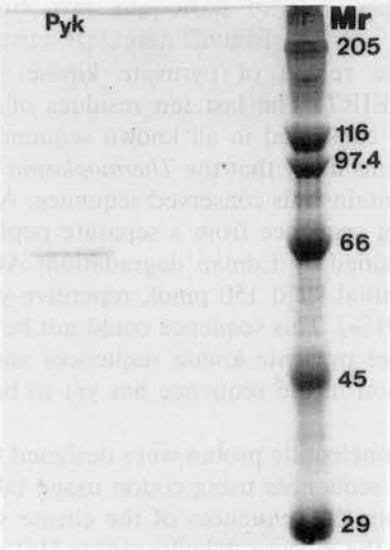


Fig. 1. SDS-PAGE gel of the purified pyruvate kinase. The gel was stained with Coomassie brilliant blue R (Sigma). The left lane contains the purified pyruvate kinase eluted from the 5′-ADP-agarose column. The right lane contains high molecular mass markers (Sigma). The M_r values shown are $\times 10^{-3}$.

Table 1

Purification table for *Thermoplasma* pyruvate kinase

Step	Activity (U)	Protein (mg)	Specific activity (U/mg)	Purifications
Cell-free extract	11.2	28.0	0.40	1
Ammonium sulphate precipitation	13.5	17.2	0.78	1.96
Mono-Q	14.0	0.56	25.0	62.5
Superose 6	13.8	0.28	49.5	124
5′-AMP-agarose	10.5	0.05	201.2	503

Units of specific activity are μmol NADH converted min^{-1} (mg of protein)$^{-1}$. Final recovery (from maximal activity found) is 75%.

The kinetic properties of the enzyme are listed in Table 2. They show that, whilst the enzyme exhibits fairly typical $K_{0.5}$ values for its various substrates and is activated by AMP, it is much more stable at higher temperatures than any of its counterparts, with the possible exception of the enzyme from *Bacillus stearothermophilus*. Indeed, previous studies [9] have shown that the enzyme is stable for over 10 min at 90°C and half an hour at 70°C. It is hoped that primary structure information gleaned from cloning studies currently in progress may shed some light on this enhanced thermostability of the *Thermoplasma* enzyme.

The amino acid composition of the pyruvate kinase has been determined and resembles that of other pyruvate kinases except that there are

the peak eluted with ethylene glycol does not react and the supernatant is uncoloured (Fig 2). When the same samples are allowed to diffuse into the agar plates containing starch, the first peak gives a clear halo indicating the presence of α- or α- and β-amylase.

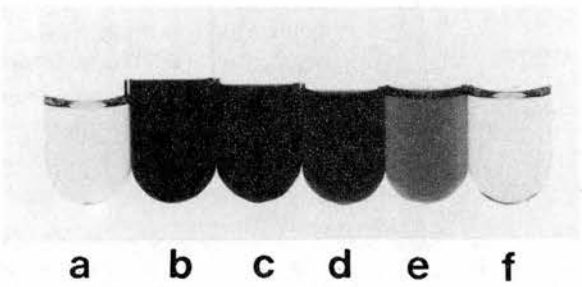


Figure 2 Differentiation of α-amylase and β-amylase activities by Phadebas amylase test. (a) blank with water; (b) malt extract; (c, d, e) different dilutions of peak I (Fig 1); (f) peak II (Fig 1)

The second peak gives a small pink halo typical from β-amylase (Fig 3). Thus, the peak eluted by ethylene glycol contains β-amylase and the peak eluted by salt deletion mainly contains α-amylase.

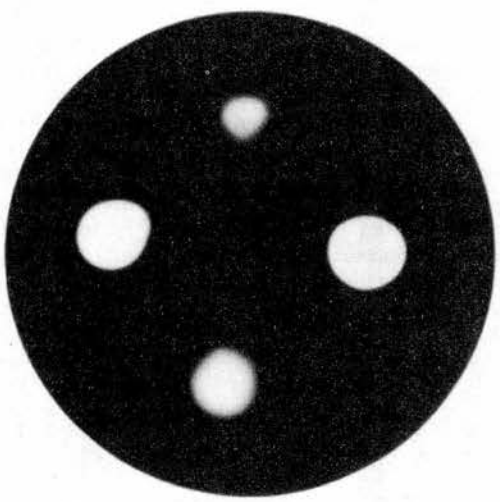


Figure 3 Differentiation of α-amylase and β-amylase activities using the agar diffusion method. Left: malt extract; right: peak I; top: sweet potato β-amylase standard; bottom: peak II.

It is important to point out that Phenyl-Sepharose beds may be re-used many times after being thoroughly washed on a sintered glass filter with distilled water, followed by ethanol and *n*-butanol. After a few minutes, the gel can be treated with these in the reverse sequence and stored in the adsorption buffer.

Conclusions

The resolution of malt extract in α- and β-amylasic fractions is of particular value as a teaching laboratory exercise because it illustrates several biochemical concepts. It shows, as expected, that the hydrophobic interactions are stronger at high ionic strength, so the adsorption to a hydrophobic adsorbent may be conveniently performed after a salt precipitation or an ion-exchange chromatography step.

The conditions used for the elution show how different parameters (hydrophobicity, polarity of the environment) influence the interaction between an amphiphilic bed and the adsorbed proteins.

The techniques used to distinguish between α- and β-amylase activities give the students an idea of two ways by which the differentiation can be carried out. Small regenerable gel beds are used allowing the saving of materials and money.

Acknowledgements

We thank IPICS, Uppsala University, Sweden and PEDECIBA (Programa para el Desarrollo de las Ciencias Básicas), Uruguay for financial support. We are indebted to Ulf Sthål for fruitful comments and Dr Jeffrey Martin for linguistic revision

References

[1] Daussant, J (1987) in 'Advances in Cereal Science and Technology', vol 9, edited by Pomeranz, Y, 47–60

[2] Friedberg, F (1985) *Biochem Educ* **13**, 105–107

[3] Bernfeld, P (1955) in *Methods in Enzymology*, vol **1**, edited by Colowick, S P and Kaplan, N O, 149–150

[4] Phadebas Amylase Test, Pharmacia Diagnostics (1983)

[5] Briggs, B A (1962) *J Inst Brew* **68**, 27–32

[6] Hjertén, S (1980) in 'Methods of Biochemical Analysis', **27**, 89–108

[7] Eriksson, K O (1989) in 'Protein Purification', edited by Janson, J C and Rydén, L, VCH, New York, pp 207–226

Molecular Evolution: The Origin of Glycolysis

SIMON POTTER and LINDA A FOTHERGILL-GILMORE

Department of Biochemistry
University of Edinburgh
George Square
Edinburgh EH8 9XD, Scotland

Introduction

Over the past decade or so, a great deal of information about how individual enzymes evolve has been gathered, largely from comparisons of protein and gene structures. These comparisons are facilitated using sequence databases which, to date, hold upwards of thirty million nucleotides of information. It is now understood, for example, that enzyme evolution occurs on various levels ranging from single amino acid changes through insertions, deletions and exon-shuffling to gene duplication/fusion and lateral gene transfer events.[1] Unfortunately, much less is known about the evolution of the metabolic pathways of which these enzymes are often a part.

A good case in point is the glycolytic pathway. This sequence of metabolic reactions is catalysed by ten enzymes and is unique in terms of the amount of information available about it. Sequences of all ten enzymes are available from a variety of sources as are their crystal structures.[2] This information has revealed some interesting interrelationships between the enzymes which may begin to shed some light on how the glycolytic pathway as a whole may have evolved. Some of the most interesting of these insights have arisen as a result of studies on the glycolytic enzymes of the archaea, a group of organisms originally distinguished by their ability to exist in extreme habitats: *methanogens*, which live in badly aerated swampy areas and the ruminant gut; *halophiles*, which live in areas of high salt content or low water activity, and *thermophiles*, which live in areas of high temperature such as hot springs. They are all prokaryotic in nature and organisation but possess macromolecular constituents which more closely resemble their eukaryotic counterparts.[3] In fact, they are now recognised as constituting a third taxonomic superkingdom or 'domain' of equal ranking with the bacteria and the eukarya (*note*: throughout this article, the new taxonomic terminology of Woese *et al*[4] is used).

Use of computers

Before discussing the information gleaned from comparative studies of archaeal glycolytic enzymes and their bacterial and eukaryal counterparts, it will be helpful to review briefly the theoretical background to the studies. If two proteins are homologous (ie are derived from a common ancestor), when their sequences are aligned they should be over 25% identical. The sequences may be aligned using computer programs such as CLUSTAL[5] which aim to produce maximum identity between sequences by introducing small gaps into them. For an example of the use of gapping in sequence alignment see Fig 1 which shows two alignments of the α and β chains of haemoglobin, one with gaps and the other without.[6] The threshold value of 25% identity is an average of the values that would be expected were two unrelated protein sequences to be aligned with gaps. The reason for this is that alignment programs allow gaps to be incorporated into either sequence if, by doing so, the alignment is significantly improved. An appropriate gap penalty is imposed in order that gapping not be unrestrained. For example, gaps are seldom introduced into regions of probable secondary structure such as α-helices. If gapping were not allowed, two random sequences would be about 5–6% identical. If two sequences show identity above the threshold value they are said to be homologous. If they show identity around the threshold value or below it, it is probable that they are not homologous (although it is impossible to say that with certainty as they may be distantly related to one another, retaining only a few vital structural and functional regions).

In the case of multiple sequence alignments, which can be performed by programs such as CLUSTAL, it is possible to construct a matrix of percentage sequence identities (or differences) which relates each protein to its counterparts from other organisms. Although more and more proteins are being subjected to this type of analysis, the pioneering work of Carl Woese and his co-workers was performed using alignments of RNA sequences rather than proteins. A matrix of rRNA differences may be used to generate a phylogenetic tree of the type shown in Fig 2. In this diagram, the branch lengths of the tree are related to the percent differences between the RNA

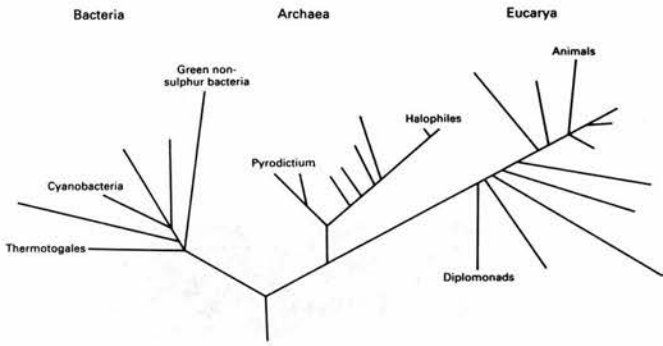


Figure 2 Rooted phylogenetic tree constructed from 16S (or 18S) rRNA sequences. Estimates of sequence divergence (mutations fixed per sequence position) were calculated and used by Woese to infer the tree

molecules of the various organisms being compared. The molecules which are used to construct these trees — usually rRNA or, increasingly often, protein molecules — are known as molecular clocks or phylogenetic markers.

The Application of Sequence Alignments to the Enzymes of Glycolysis

Studies on the 3-phosphoglycerate kinases (PGK) of methanogenic archaea have shown these proteins to be 30–36% identical to their bacterial and eukaryal counterparts.[7] It is probable, therefore, that PGKs are all descended from a common ancestor. This is complemented by a preliminary study of the pyruvate kinase (PYK) of a thermophilic archaeum[8] which also shows considerable identity to regions of other PYKs. These data suggest a phylogenetic tree similar to that shown in Fig 2. In contrast, it appears that the archaeal glyceraldehyde-3-phosphate dehydrogenases (GAPDH) are not homologous with other GAPDHs as they exhibit only 15% identity to them.[9] This suggests that the archaeal GAPDH is descended from an entirely different protein than that which gave rise to the bacterial

Figure 1 Alignments of α- and β-haemoglobins with introduced gaps (upper panel) or with direct sequence alignment without gaps (lower panel). The gaps in the upper panel were inserted at three places (arrows) to maintain obviously similar sequences in register

karyal GAPDHs. It is, therefore, possible that glycolysis arose an association of enzymes which worked on substrates of nilar structure. Some of the enzymes therefore share common cestors and some of which are descended from different arent' molecules. This information was recently exploited by e authors as a basis for an examination problem for final year dergraduate students (see sample problem section).

ycolysis in the Thermophilic Archaea

nother valuable insight into the origins of glycolysis has come om the discovery that this pathway is not ubiquitous as had ce been thought. In fact, in some thermophilic archaea, only o of the common ten enzymes are present, enolase (ENO) and YK.[10] This may suggest that the glycolytic pathway may have olved from the 'bottom up' (ie in the gluconeogenic sense) arting with ENO and PYK, the only two parts of the pathway ll thought to be ubiquitous. As time progressed, other zymes joined the pathway some of which became common to organisms. The final event in the assembly of the eukaryal ycolytic pathway was presumably the incorporation of 6-osphofructokinase, the point at which bacterial glycolysis is ocked. Some of the thermophilic archaea seem to have been t behind by this process and hence may provide a glimpse into e distant evolutionary past of glycolysis.

Glycolysis may not be the only pathway to have its distant igins in thermophilic archaea. *Archaeglobus fulgidus*, a sulph-e reducer, has recently been shown to possess elements of the ethanogenerative pathway which was once thought to be ique to the methanogens.[11] In conclusion, our opinions about e evolution metabolic pathways are still largely based on eculation but, with the recognition of the archaea and their usual biochemistry, a powerful window onto the past has been ened by the molecular evolutionist.

mple Problem for a Molecular Evolution Examination

ne following problem is based on work by Hensel *et al*[9] on the APDH enzymes of archaea and their relatedness to GAPDHs om organisms from the other two domains. The students were st given general information about GAPDH which catalyses e oxidative phosphorylation of glyceraldehyde 3-phosphate to ycerate 1,3-bisphosphate. In most organisms the enzyme has ur identical subunits each of 330–340 amino acid residues, and e cofactor required for the reaction is NAD^+. The amino-rminal half of the enzyme is involved in cofactor binding and e carboxy-terminal half provides the residues directly involved catalysis. The students were also provided with information out the archaeal GAPDH: an enzyme recently isolated from e archaeum *Methanobacterium formicicum* that catalyses the me reaction except that the preferred cofactor is $NADP^+$. The zyme is a homotetramer with 337 amino acid residues per bunit.

The students were provided with a sequence alignment of APDHs from five sources which had been produced with LUSTAL. The sequences aligned were from *Drosophila*, man, yeast, *Bacillus stearothermophilus* and *Methanobac-ium formicicum*. The alignment is shown in Fig 3 (although dents were given an alignment of only the carboxy-terminal lves of the enzymes, from the residue marked with the *). ven this alignment, the students were asked to:

) construct a table (matrix) of percent amino acid residue ferences from all pairwise comparisons of the sequences.

) use the information from the matrix to construct a simple ylogenetic tree.

e students were then asked to discuss the possible evolution-history of the enzyme and to assess the usefulness of APDH as a molecular clock.

A matrix of differences is shown in Table 1 and the ylogenetic tree constructed from it is shown in Fig 4. The ntral importance attached to the tree was that it should

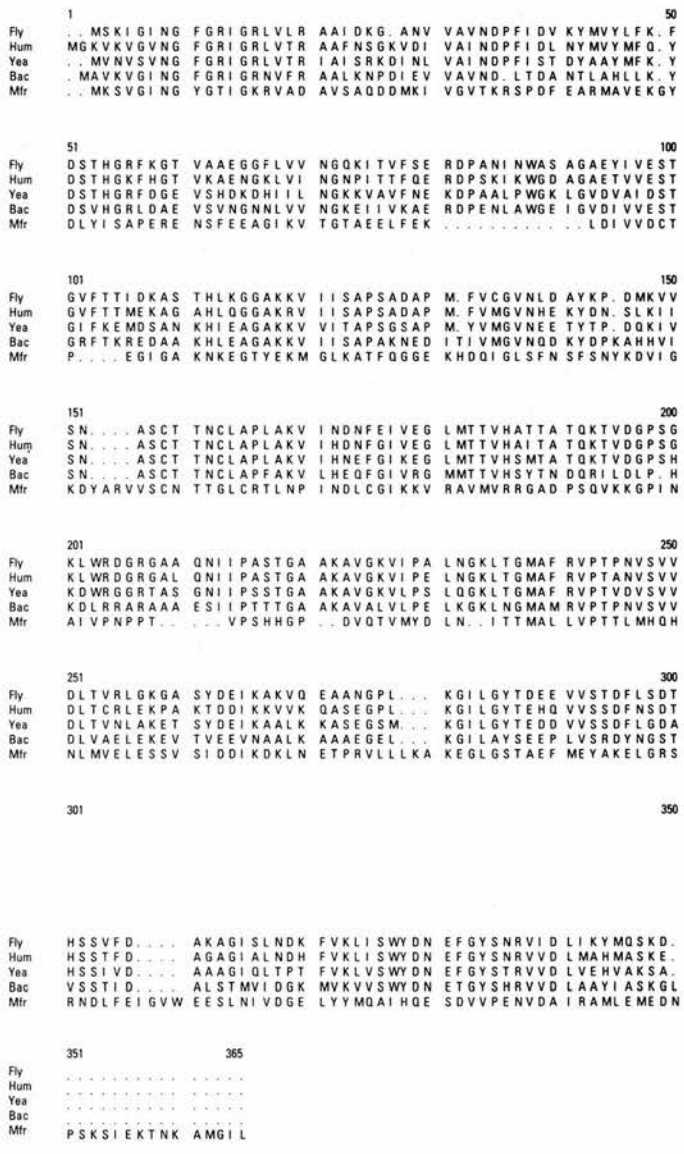


*Figure 3 Sequence alignment of glyceraldehyde 3-phosphate dehydrogenase enzymes. Fly (*Drosophila*), Hum (human), Yea (*Saccharomyces cerevisiae*), Bac (*Bacillus stearothermophilus*), Mfr (*Methanobacterium formicicum*). The alignment was produced using the program CLUSTAL and the sequences were obtained from the PIR database using the UWGCG package*

Table 1 Matrix of differences between GAPDHs. Key as Figure 3

	Fly	Hum	Yea	Bac	Mfr
Fly	—				
Hum	30	—			
Yea	39	37	—		
Bac	49	51	48	—	
Mfr	86	89	89	88	—

correctly reflect the grouping of the different organisms. Hence, an approximation of the tree shown in Fig 4, showing the yeast, *Drosophila* and human sequences clustered together, the *Bacillus* sequence separated by a short branch, and the archaeal sequence separated by a much longer branch was all that was required from the students. In the essay part of the paper, the students were asked to deduce from their crude tree that the archaeal enzyme is probably unrelated to the enzymes from the

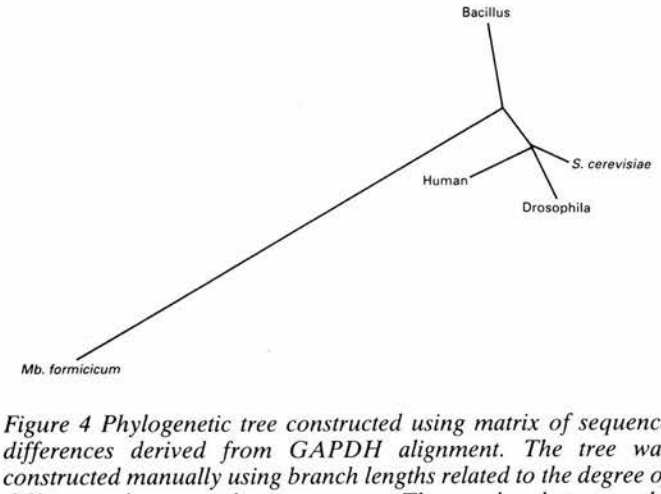


Figure 4 Phylogenetic tree constructed using matrix of sequence differences derived from GAPDH alignment. The tree was constructed manually using branch lengths related to the degree of difference between the sequences. The angles between the branches are arbitrary, assigned simply to make the viewing of the diagram easier

bacterial and eucaryal sources which are interrelated. From this observation, it is clear that GAPDH would not be a useful molecular clock for comparing organisms from all three domains although it could be used for bacterial and eucaryal trees.

References

[1] Fothergill-Gilmore, L A (1991) 'The evolution of RNA and proteins as biocatalysts,' in *Fundamentals of Medical Cell Biology*, by Bittar, E E (editor) Vol **1**, *Evolutionary biology*, JAI Press, London, pp 163–188

[2] Fothergill-Gilmore, L A and Michels, P A M (1992) 'Evolution of glycolysis,' *Prog Biophys Mol Biol (in the press)*

[3] Rivera, M C and Lake, J A (1992) *Science* **257**, 74–76

[4] Woese, C R, Kandler, O and Wheelis, M L (1990) *Proc Natl Acad Sci, USA* **87**, 4576–4579

[5] Higgins, D G and Sharp, P M (1988) *Gene* **73**, 237–244

[6] Doolittle, R F (1986) 'Of Urfs and Orfs: A Primer on How to Analyse Derived Amino Acid Sequences,' University Science Books, Mill Valley, California

[7] Fabry, S, Heppner, P, Dietmaier, W and Hensel, R (1990) *Gene* **91**, 19–25

[8] Potter, S and Fothergill-Gilmore, L A (1992) *FEMS Microbiol Lett* **94**, 235–240

[9] Hensel, R, Zwickl, P, Fabry, S, Lang, J and Palm, P (1989) *Can J Microbiol* **35**, 81–85

[10] Danson, M J (1988) *Adv Microb Physiol* **29**, 165–231

[11] Moller-Zinkham, D *et al* (1989) *Arch Microbiol* **152**, 362–367

Book Reviews

1992 Supplement to Biochemistry

Donald Voet and Judith G Voet. pp 73. John Wiley and Sons, New York. 1992. $2.95/£1.95 0–471–57944–0

Is it worth while to review a slim volume that costs only a nominal sum? In the present case, the novel objective and unusual character of the book raise questions of general interest to teachers of biochemistry that merit analysis. The authors of a successful textbook are to be commended for seeking a novel solution to the problem of preventing their efforts from suffering rapid obsolescence. I wish that I could conclude that they hav succeeded.

The first question concerns the technical problem of integra ing new information so that readers are necessarily aware additions and alterations. There might be several ways approach this problem, including an index to the supplemen references to specific paragraphs in the original text, or a list relationships. Since none of these exists, it is difficult to imagir how they might work and I am left with the feeling that studen are not likely to integrate supplementary material easily as the use the text either as a primary source of information or as reference.

In the introduction to the second supplement to the textbook, Donald and Judith Voet emphasize the importance students and teachers keeping up with the literature. Howeve they do not make it clear whether the supplement is intende primarily for new students, who should be introduced to ne information and recent articles as part of their first course biochemistry, or whether it is intended for students who hav completed their courses and are encouraged to update their tex so that the familiar book need not be replaced by more mode editions. Attempts by this reviewer to relate the supplementa material to the original text leave considerable doubt wheth either purpose is served. Although I agree with the importan of students and practitioners of biochemistry surveying th literature regularly and familiarizing themselves with releva items, I believe that the selection of relevant literature is high personal and must be done individually as a function professional interests. Therefore, my comments deal only wi the supplement as a manual to be used together with th textbook during an introductory course in biochemistry.

It should be noted that the supplement contains bri discussions of information from current literature but that mu of the supplement consists of mere references to review articl and primary papers. The 90 items selected for discussion avera less than 3 per chapter; no additional text is given for 10 of th chapters and 11 of the items are related to the chapter eukaryotic gene expression. In addition to this material, 1 references are listed as suggested reading for 25 of the chapters. Do these references selected from perhaps 10 0 articles that appear each year in major biochemical journa represent basic information that would be likely to be included a new text? The examples below suggest that the selecti reflects the interests of the authors in structure-function relatio ships and that the items that have caught their fancies a presented in much greater detail than similar items are discuss in the original text.

Striking evidence of the personal interests of the authors is t fact that 40 of the 90 topics elaborated in the supplement a based on X-ray crystallography. There is no question of t value of this technology to elucidation of many problems of gre current interest to active investigators in many areas biochemistry. It would be a gross exaggeration, however, maintain the enthusiasm that greeted David Phillips' study lysozyme with the belief that the power of X-rays had ma other aspects fo enzymology obsolete. One paper that included deals with the fact that the core protein of the Sindl virus is a serine protease. Since the virus is not described either the original text or the supplement, the context of t information is obscure. It is also not clear why this seri protease is particularly worthy of addition to the 13 diver members listed in the original text.

The only item presented as an advance in amino ac metabolism is an elaborate discussion of the mechanism which chloroquine acts to kill the plasmodia that cause malar Although the studies do present new information in the w these organisms metabolize heme, it is doubtful that such unique reaction is typical of the sort of information that shou be included in any textbook.

Johnson, J.L., and Francis, B.S. 1975. Taxonomy of the *Clostridia*: ribosomal ribonucleic acid homologies among the species. J. Gen. Microbiol. 88: 229–244.

Leirmo, S., Harrison, C., Cayley, D.S., Burgess, R.R., and Record, M.T. 1987. Replacement of potassium chloride by potassium glutamate dramatically enhances protein-DNA interactions *in vitro*. Biochemistry, 26: 2095–2101.

MacDonald, R.E., and MacDonald, S.W. 1962. The physiology and natural relationships of the motile, sporeforming *Sarcinae*. Can. J. Microbiol. 8: 795–808.

Magill, N.G., Loshon, C.A., and Setlow, P. 1990. Small, acid-soluble, spore proteins and their genes from two species of *Sporosarcina*. FEMS Microbiol. Lett. 72: 293–298.

Measures, J.C. 1975. Role of amino acids in osmoregulation in non-halophilic bacteria. Nature (London), 257: 398–400.

Murrell, W. 1969. Chemical composition of spores and spore structures. *In* The bacterial spore. *Edited by* G.W. Gould and A. Hurst. Academic Press, London. pp. 215–273.

Nelson, D.L., and Kornberg, A. 1970*a*. Biochemical studies of bacterial sporulation and germination. XVIII. Free amino acids in spores. J. Biol. Chem. 245: 1101–1107.

Nelson, D.L., and Kornberg, A. 1970*b*. Biochemical studies of bacterial sporulation and germination. XIX. Phosphate metabolism during sporulation. J. Biol. Chem. 245: 1137–1145.

Nelson, D.L., Spudich, J.A., Bonsen, P.P.M., Bertsch, L.L., and Kornberg, A. 1969. Biochemical studies of bacterial sporulation and germination. XVI. Small molecules in spores. *In* Spores IV. *Edited by* L.L. Campbell. American Society for Microbiology, Washington, D.C. pp. 59–71.

Pechman, K.J., Lewis, B.J., and Woese, C.R. 1976. Phylogenetic status of *Sporosarcina ureae*. Int. J. Syst. Bacteriol. 26: 305–310.

Rotman, V., and Fields, M.L. 1967. A modified reagent for dipicolinic acid analysis. Anal. Biochem. 22: 168.

Setlow, P., and Kornberg, A. 1970. Biochemical studies of bacterial sporulation and germination. XXII. Energy metabolism in early stages of germination of *Bacillus megaterium* spores. J. Biol. Chem. 245: 3637–3644.

Setlow, P., and Primus, G. 1975. Protein metabolism during germination of *Bacillus megaterium* spores. I. Protein synthesis and amino acid metabolism. J. Biol. Chem. 250: 623–630.

Slepecky, R., and Foster, J.W. 1959. Alterations in the metal content of spores of *Bacillus megaterium* and the effect on some spore properties. J. Bacteriol. 78: 117–123.

Stackebrandt, E., Ludwig, W., Weizenegger, M., Dorn, S., McGill, T.J., Fox, G.E., Woese, C.R., Schubert, W., and Schleifer, K.-H. 1987. Comparative 16S rRNA oligonucleotide analyses and murein types of round-spore-forming *Bacilli* and non-spore-forming relatives. J. Gen. Microbiol. 133: 2523–2529.

Stewart, M., Somlyo, A.P., Somlyo, A.V., Shuman, H., Lindsay, J.A., and Murrell, W.G. 1980. Distribution of calcium and other elements in cryosectioned *Bacillus cereus* T spores, determined by high-resolution scanning electron probe X-ray microanalysis. J. Bacteriol. 143: 481–491.

Swerdlow, B.M., Setlow, B., and Setlow, P. 1981. Levels of H^+ and other monovalent cations in dormant and germinating spores of *Bacillus megaterium*. J. Bacteriol. 148: 20–29.

Thompson, R.S., and Leadbetter, E.R. 1963. On the isolation of dipicolinic acid from endospores of *Sarcina ureae*. Arch. Mikrobiol. 45: 27–32.

Warth, A.D. 1978. Molecular structure of the bacterial spore. Adv. Microb. Physiol. 17: 1–45.

Evidence for a dual-specificity isocitrate dehydrogenase in the euryarchaeotan *Thermoplasma acidophilum*

SIMON POTTER

Department of Biochemistry, University of Edinburgh, Hugh Robson Building, George Square, Edinburgh EH8 9XD, United Kingdom

Received July 13, 1992

Revision received September 23, 1992

Accepted September 28, 1992

POTTER, S. 1993. Evidence for a dual-specificity isocitrate dehydrogenase in the euryarchaeotan *Thermoplasma acidophilum*. Can. J. Microbiol. 39: 262–264.

The thermoacidophilic euryarchaeotan *Thermoplasma acidophilum* occupies a paradoxical place in phylogenetic trees. It has been found to possess both NAD- and NADP-dependent isocitrate dehydrogenase activities. Kinetic evidence presented suggests that both activities are functions of the same protein. The significance of this result, evolutionary and otherwise, is discussed.

Key words: archaea, isocitrate dehydrogenase, molecular evolution.

POTTER, S. 1993. Evidence for a dual-specificity isocitrate dehydrogenase in the euryarchaeotan *Thermoplasma acidophilum*. Can. J. Microbiol. 39 : 262–264.

Thermoplasma acidophilum, une euryarchébactérie thermophile, occupe une position paradoxale dans les arbres phylogénétiques. Cette bactérie possède les deux activités NAD- et NADP-dépendantes de l'isocitrate déshydrogénase. Les résultats cinétiques présentés suggèrent que les deux activités sont reliées à la même protéine. On discute de la signification de ce résultat, sur le plan de l'évolution ou autre.

Mots clés : archaea, isocitrate déshydrogénase, évolution moléculaire.

[Traduit par la rédaction]

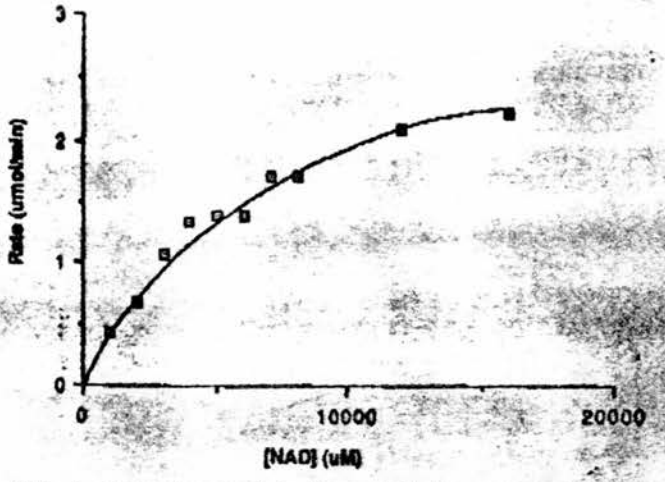


FIG. 1. Dependence of isocitrate dehydrogenase activity on [NAD^+]. The assay was performed as described in the text.

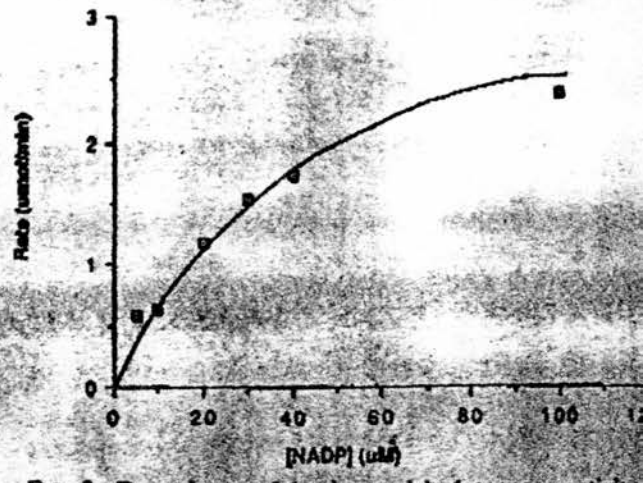


FIG. 2. Dependence of isocitrate dehydrogenase activity [$NADP^+$]. The assay was performed as described in the tex

Isocitrate dehydrogenase catalyses the oxidative decarboxylation of isocitrate to 2-oxoglutarate. Eukarya contain two activities of this enzyme: an NAD-linked enzyme, confined to the mitochondria, which is allosterically regulated and has a controlling role in energy production by the citric acid cycle, and an NADP-linked enzyme, which is not allosterically regulated. Most bacteria have only the NADP-specific, nonallosteric enzyme (Weitzman 1981). Danson and Wood (1984) reported finding both isocitrate dehydrogenase activities present in the sulphur-dependent, thermoacidophilic chrenarchaeotan *Sulfolobus acidocaldarius*. Further to this, evidence was presented to suggest that, unlike in eukarya, both NAD- and NADP-dependent activities were functions of the same enzymic protein (Danson and Wood 1984).

Thermoplasma acidophilum, a representative organism from another branch of the archaeal tree, occupies a paradoxical place in current phylogenies (Klenk *et al.* 1986). Phenotypically, it is a chrenarchaeotan growing optimally at pH 2 and 59°C. Phylogenetically, however, it classifies with the euryarchaeota. It is possible that *T. acidophilum* may represent a 'missing link' between two branches of the archaeal tree and hence, studies of the proteins of this unique organism may have profound implications for molecular evolution hypotheses.

It is reported here that *T. acidophilum* also possesses both isocitrate dehydrogenase activities and that kinetic studies suggest that, like *S. acidocaldarius*, both activities are functions of the same protein molecule.

Cell culture and disruption

Thermoplasma acidophilum was supplied as a freeze-dried culture by the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany (DSM 1726). Cells were cultured in 2-L batches at 59°C and pH 2.0 in the medium described in the DSM catalogue of strains, which was both aerated and agitated by pumping air into it using a Pharmacia P-3 peristaltic pump. The extreme growth conditions precluded the possibility of any atmospherically carried bacterial contamination. After a 50-h culture period, the cells had reached stationary phase with an OD 650 nm of 0.26. These cells (0.2 g wet weight) were collected by centrifugation and resuspended in 1.5 mL of lysis buffer (20 mM NaCl, 10 mM $MgCl_2$, 10 mM Tris-HCl (pH 7.5), containing 1,10-phenanthroline, *N*-(*N*-(L-3-transcarboxyoxirane-2-carbonyl)-L-leucyl)-agmat and 3,4-dichloroisocoumarin protease inhibitors, all 10 m and all supplied by Boehringer Mannheim). This cell susp sion was homogenised with glass beads for 5 min at ro temperature and cell debris was removed by brief centrifu tion in a microcentrifuge.

Isocitrate dehydrogenase assay

Isocitrate dehydrogenase was assayed at 59°C in the cr extract prepared as above, which typically contained 28 of protein. The assay buffer contained 20 mM Tris-H (pH 8.0), 1 mM EDTA, 10 mM $MgCl_2$, 0.2 mM NAD (or 1.0 mM NAD^+), and 4.0 mM isocitrate substr (Sigma). The reaction, in a final volume of 1 mL, started with the addition of the enzyme and the increas A_{340} was followed with time. In crude extracts *T. acidophilum*, both NAD- and NADP-linked isocitr dehydrogenase activities were detected. The activi depended hyperbolically on the concentrations of the cof tors (NAD^+ K_m^{app} = 6.6 ± 0.6 mM; $NADP^+$ K_m^{app} 32 ± 4 μM) (Figs. 1 and 2). For each cofactor the V was similar (3.23 ± 0.03 μmol/min for the NAD-linl activity and 3.14 ± 0.04 μmol/min for the NADP-linl activity). These values are broadly in agreement with th found for the enzyme isolated from *Sulfolo acidocaldarius* (Danson and Wood 1984).

Evidence for a single isocitrate dehydrogenase with d cofactor specificity

Assays of isocitrate dehydrogenase in extracts *T. acidophilum* were carried out using NAD^+, NADP and NAD^+ plus $NADP^+$ (Table 1). The activities no for each cofactor were not additive, which indicates that t are not the result of two independent activities. In fact, results suggest that the two cofactors compete with another, as was found to be the case with the *S. aci caldarius* activities, and do so with K_i values approximat equal to their K_m values, an observation that suggests t the cofactors bind at the same site on the enzyme. T theoretical enzyme activity for one enzyme capable of us both cofactors was calculated according to the equati (Segel 1975)

$$V_{total} = \frac{V_{NAD}(1 + [NAD]/K_m^{NAD}) + V_{NADP}(1 + [NADP]/K_m^{NA}}{1 + [NAD]/K_m^{NAD} + [NADP]/K_m^{NADP}}$$

TABLE 1. Substrate competition between cofactors for *Thermoplasma acidophilum* isocitrate dehydrogenase

Substrate	Activity (μmol product·min^{-1}·mL)$^{-1}$		Theoretical activity
	T. acidophilum	*S. acidocaldarius*[a]	
1.0 mM NAD^+	0.17	0.19	
0.2 mM $NADP^+$	0.94	0.94	
1.0 mM NAD^+ + 0.2 mM $NADP^+$	0.90	0.91	0.94

NOTE: Theoretical activity for one enzyme capable of using both cofactors was calculated according to Segel (1975).
[a]Taken from Danson and Wood (1984).

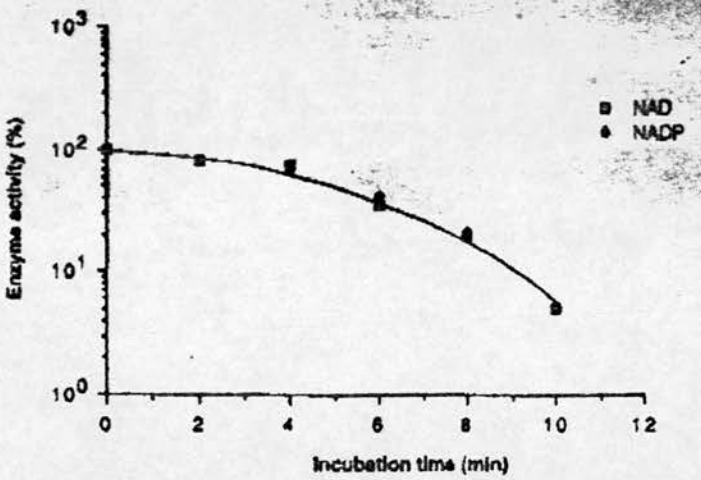


FIG. 3. Thermal inactivation of isocitrate dehydrogenase. Crude extracts of *Thermoplasma acidophilum* were treated as described in the text and the resulting activity was tested with both cofactors at known time intervals.

V_{NAD} and V_{NADP} are enzyme velocities in NAD^+ or $NADP^+$ alone and the K_m values are as above.

Thermal inactivation of the isocitrate dehydrogenase

Crude extracts of *T. acidophilum* were incubated at 90°C and samples were removed at known time intervals, allowed to cool to the assay temperature, and then assayed using the procedure described above. The results of this experiment, illustrated in Fig. 3, support the conclusion that a single enzyme molecule is involved in both activities: they are lost coincidentally.

The kinetic evidence presented here indicates that, like *S. acidocaldarius*, *T. acidophilum* also possesses an isocitrate dehydrogenase that can utilise both NAD^+ and $NADP^+$. (Since the K_m for NADH is 200-fold higher than that for NADPH, it is possible that the use of NADH by the isocitrate dehydrogenase is not physiologically significant; however, many enzymes have been observed with weaker NADH–NADPH specificities.) As the two organisms are from different branches of the archaeal tree, this further suggests that the phenomenon may be widespread throughout the archaea. This is in contrast with the bacteria, which in most cases have only the NADP—linked isocitrate dehydrogenase, and the eukarya, which have both activities but as functions of separate enzymes. Taken together, these results may help to confirm the status of the archaea as a separate taxonomic 'domain' of organisms (Woese *et al.* 1990) and to refute other work that splits up the archaea, combining some with eukarya and others with bacteria (Lake 1989).

Complementary to this work, several other groups have observed archaeal proteins that show dual specificity towards NAD^+ and $NADP^+$ cofactors: glyceraldehyde 3-phosphate dehydrogenase (Fabry and Hensel 1987), glucose dehydrogenase (Danson 1988), and malate dehydrogenase (Grossebuter *et al.* 1986). Whilst this phenomenon is certainly not specific to the archaea, it appears to be rather more prevalent amongst their enzymes than the comparable enzymes from both the bacterial and eucaryal domains. It is, therefore, possible that the detection of such dual-specificity enzymes may assist in the classification of newly discovered organisms (Fuhrman *et al.* 1992) in conjunction with other characteristics such as the lipid composition of plasma membranes.

Acknowledgements

Thanks are due to Dr. Linda Fothergill-Gilmore for her patient supervision and timely advice. The financial support of the Science and Engineering Research Council is gratefully acknowledged.

Danson, M.J. 1988. Central metabolic pathways of archaebacteria. Adv. Microb. Physiol. 29: 165–231.

Danson, M.J., and Wood, P.A. 1984. Isocitrate dehydrogenase of the thermoacidophilic archaebacterium *Sulpholobus acidocaldarius*. FEBS Lett. 172: 289–293.

Fabry, S., and Hensel, R. 1987. Purification and characterisation of D-glyceraldehyde 3-phosphate dehydrogenase from the thermophilic archaebacterium *Methanothermus fervidus*. Eur. J. Biochem. 165: 147–155.

Fuhrman, J.A., McCallum, K., and Davis, A.A. 1992. Novel major archaebacterial group from marine plankton. Nature (London), 356: 148–149.

Grossebuter, W., Hartl, T., Gorisch, H., and Stezowski, J.J. 1986. Purification and properties of malate dehydrogenase from the thermoacidophilic archaebacterium *Thermoplasma acidophilum*. Biol. Chem. Hoppe-Seyler, 367: 457–463.

Klenk, H.P., Haas, B., Schwass, V., and Zillig, W. 1986. Hybridisation homology—a new parameter for the analysis of phylogenetic relations demonstrated with the urkingdom of the archaebacteria. J. Mol. Evol. 24: 167–173.

Lake, J.A. 1989. Origin of the eukaryotic nucleus: eukaryotes and eocytes are genotypically related. Can. J. Microbiol. 35: 109–118.

Segel, I.H. 1975. Analysis of the steady state kinetics of multi-substrate systems. *In* Enzyme kinetics: behaviour and analysis of rapid equilibrium and steady state systems. Wiley, New York. pp. 113–118.

Weitzman, P.D.J. 1981. Unity and diversity in some bacterial citric acid cycle enzymes. Adv. Microb. Physiol. 22: 185–244.

Woese, C.R., Kandler, O., and Wheelis, M.L. 1990. Towards a natural system of organisms: proposal for the domains archaea, bacteria and eucarya. Proc. Natl. Acad. Sci. U.S.A. 87: 4576–4579.

The pyruvate kinase of *Thermoplasma acidophilum*: purification, kinetic characterisation & use as a phylogenetic marker.

SIMON POTTER and LINDA A. FOTHERGILL-GILMORE

Department of Biochemistry, University of Edinburgh, George Square, Edinburgh EH8 9XD. U.K.

Glycolysis is a central catabolic pathway involving 10 enzymes which was, until recently, assumed to be ubiquitous. This assumption was put into doubt by the discovery that the archaebacteria have varying proportions of the pathway's enzymes [1]. More specifically, the thermoacidophilic archaebacterium *Thermoplasma acidophilum* is known to possess operational glycolysis [2] whilst having none of the following enzyme activities: 6-phosphofructokinase, fructose-1,6-bisphosphate aldolase, glyceraldehyde-3-phosphate dehydrogenase and phosphoglycerate mutase [3].

The archaebacteria then may provide a unique glimpse into the evolution not only of the enzymes of glycolysis, but also of the pathway as a whole. Generally, comparisons of the sequences of an enzyme isolated from different organisms may help explain its evolution and, specifically, such studies have shown that the glycolytic enzymes are particularly highly conserved with only a 5% residue change per 100 million years [4].

The aim of this project is to isolate and sequence the gene of an archaebacterial glycolytic enzyme for the purposes of characterisation and comparison with its mesophilic counterparts. This will then be a useful addition to the information used to construct phylogenetic trees from sequence data. The enzyme of choice was Pyruvate kinase (ATP-pyruvate 2-O-phosphotransferase, EC 2.7.1.40) which catalyses the transphosphorylation from phosphoenolpyruvate (PEP) to ADP and requires Mg^{2+} and K^{+} .This enzyme was chosen because it has been found in all organisms tested for its presence and retains the same function in all of them, hence fulfilling the requirements of a phylogenetic marker.

The pyruvate kinase was purified to homogeneity on an SDS-polyacrylamide gel stained with Coomassie brilliant blue using a combination of ammonium sulphate precipitation and size-exclusion, ion-exchange and affinity chromatography techniques. The enzyme is thought to be a tetramer of molecular weight 250000 (as determined by Superose-12 size exclusion chromatography) with four identical subunits of 64000 molecular weight (as determined by SDS-polyacrylamide gel electrophoresis).

The enzyme has a final specific activity of 200U/mg as determined by the coupled assay of Bucher and Pfleiderer [5] and shows typical Km values for the substrates ADP and PEP. Further, it can be controlled by a number of effector molecules such as AMP and ATP. It has, however, significantly greater stability at higher temperatures than mesophilic pyruvate kinases as shown in figure 1.

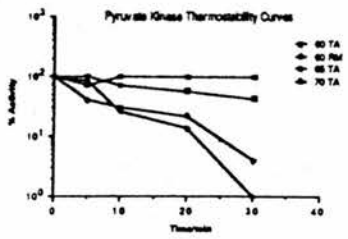


Fig.1 *Thermostability of the pyruvate kinase compared to that of the rabbit muscle enzyme.* The *Thermoplasma* enzyme (TA) was incubated at 60, 65 and 70°C respectively whilst the rabbit muscle enzyme (RM) was incubated at 60°C. After incubation, the enzymes were allowed to cool to 45°C and were then assayed according to the lactate dehydrogenase coupled assay [5].

Preliminary studies demonstrated that the N-terminus of the enzyme is blocked and hence the enzyme was cleaved into peptides in order to obtain internal sequence. The strategy was as follows: the protein was first succinylated with succinic anhydride (to prevent tryptic cleavage at lysine residues); it was then digested with either clostripain or TPCK-treated trypsin; peptides were separated on a Pharmacia f.p.l.c. system using the PEP-RPC column with a linear gradient of 0-30% acetonitrile; peaks were collected and tested for purity on an Applied Biosystems 130A h.p.l.c system using a microbore reverse phase column; approximately 20pmol of each peptide was submitted for quantitation by amino-acid analysis; selected peptides were then sequenced.

Two stretches of internal sequence have so far been obtained: a conserved region - ALDTKGPEIRT and a non-conserved region - AGIYLPGAN. Oligonucleotide probes are being made for these regions using codon usage tables derived from the genes so far sequenced from *T.acidophilum* and will be used for Southern blotting with restriction digests of *T.acidophilum* genomic DNA.

The financial support of the Science and Engineering Research Council is gratefully acknowledged.

1. Danson M. J. (1988) *Adv. Microb. Phys.* **29** 165-223
2. Searcy D. G. & Whatley F. R. (1984) *System. Appl. Microb.* 5 30
3. Budgen N. & Danson M. J. (1986) *FEBS lett.* **196** 207-210
4. Fothergill-Gilmore L. A. (1987) *Biochem. Soc. Trans.* 15 993-995
5. Bucher S. & Pfleiderer M. L. (1955) *Meth. Enzymol.* 1 47-52

Purification and partial characterisation of a protein of unknown function from *Thermoplasma acidophilum*

SIMON POTTER, NICK JOHNSTON and LINDA A. FOTHERGILL-GILMORE

Department of biochemistry, University of Edinburgh, George Square, Edinburgh EH8 9XD, U.K.

The archaebacteria have recently been proposed to be a third, distinct lineage of organisms, as different from the true bacteria as they are from eukaryotes. To relate the organisms of these three 'domains' to one another, phylogenetic trees are constructed using data from amino acid and nucleotide sequence databases [1]. *T. acidophilum* is a thermoacidophilic archaebacterium which thrives at 59°C and pH 2. A protein of unknown function has been purified and partially characterised from this organism. The protein, which has an Mr of 20K (determined by SDS-PAGE) was purified to homogeneity by a combination of ammonium sulphate precipitation, ion-exchange chromatography and size exclusion chromatography techniques. Both chromatographic steps were performed on the Pharmacia fast protein liquid chromatography system. An SDS-PAGE gel of the purified protein is shown in fig. 1.

Further to this, in an attempt to identify the protein, its amino acid composition and N-terminal sequence have been determined. The amino acid composition is shown in table 1. The first twenty amino acids of the N-terminus of the protein were determined by automated Edman degradation using an Applied Biosystems 477A

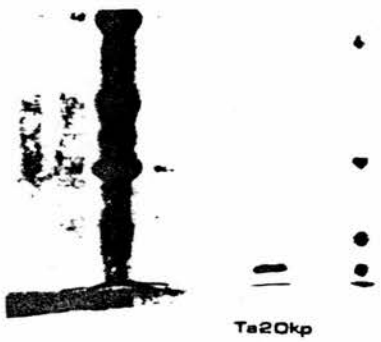


Figure 1: SDS-PAGE of purified 20K protein. L-R, high Mr markers, 20K protein, low Mr markers.

Table 1: Amino acid composition of 20K protein

D/N 20	A 36	F 1
E/Q 24	P 7	K 8
S 12	Y 1	M 1.5
G 21	V 15.5	H 1.5
R 2	I 18	T 10
L 9	TOTAL 187.5	

Cys and Trp not determined

instrument with a 120A on-line phenylthiohydantoin analyser [2]. The sequence obtained from an analysis of 250 pmol of the protein was: **AVKVGDKAPDFEAPDTNLKM.** Repetitive yield for the alanines was 92% for an intial yield of 243 pmol. The sequence was fed into the NBRF and SWISSPROT protein sequence databases and several possible matches were found, amongst them subtilisin and the M protein from *Streptococcus* species. The protein also shows an intriguing similarity to glyceraldehyde 3-phosphate dehydrogenases of archaebacteria although physically it is much smaller.

The future aims of the project are twofold. First, more sequence is to be obtained. This is to be acheived primarily by fragmenting the protein with a proteolytic enzyme such as clostripain and gaining internal sequence from the resultant peptides (as has been done on the pyruvate kinase of *T. acidophilum* [3]). Secondly, the N-terminal sequence has been used to design an oligonucleotide probe using a codon usage table derived from the citrate synthase sequence [4] of *T. acidophilum*. This probe has been successfully used to identify fragments of *T. acidophilum* genomic DNA for cloning experiments. Once the protein has been identified and its complete sequence determined, it is hoped that this information can be added to that used in the construction of phylogenetic trees.

[1] Woese, C.R. *et al.* (1990) Proc. Natl. Acad. Sci. U.S.A. 87, 4576-4579

[2] Hayes, J.D. *et al.* (1989) Biochem. J. 264, 437-445

[3] Potter, S. and Fothergill-Gilmore, L.A. (1992) FEMS Microbiol. lett. 94, in press.

[4] Sutherland, K.J. *et al.* (1991) FEBS lett. 282, 132-134

Through Semantics to Systematics:
The Trouble with Archaebacteria

WILLIAM Shakespeare once said that "A rose by any other name would smell as sweet". I doubt if today's taxonomists would agree with him. Taxonomy in general (and its naming function specifically) is currently undergoing one of its occasional upheavals. The spark to this particular fire has been the 'discovery' of the archaebacteria, a branch of organisms which is purported in some quarters to be a third, totally separate 'domain' of the tree of life (N.B. a glossary of terms herein is provided in Table 1). The resultant controversy has polarised the interested scientific community into essentially two camps which I shall term the 'radicals' and the 'conservatives'.

The Radicals

The radicals are the broader church, comprising several (very politely) warring factions. They all propose new phylogenies based upon evidence gathered from comparative studies of related macromolecular sequences. These molecules are termed 'phylogenetic markers' and should fill the following criteria: first they should have recognisable counterparts in all organisms, and secondly they should maintain a constancy of function (this will ensure that any evolutionary changes are selectively neutral). Each of the factions, however, uses different sequences as their markers and draws different conclusions from them. I shall briefly refer to three of these factions.

The first group is led by Carl Woese and bases its studies on ribosomal RNA molecules (rRNA). Bacterial small subunit rRNA is distinguishable from that of both eukaryotes and archaebacteria by the hairpin loop between positions 500 and 548 of the molecule. The loop has a side bulge protuding from the upstream strand of the stalk which, in all bacteria studied so far, comprises seven nucleotides and protudes from between the sixth and seventh base pairs of the stalk. The archaebacterial rRNA meanwhile possesses a unique structure between positions 180 and 197 which distinguishes it from the bacterial eukaryotic versions of the molecule. The upshot of this work is the hypothesis that the archaebacteria form a unique, monophyletic (holophyletic) lineage of their own and are one of three 'domains' (a kind of superkingdom), the others being eukaryotes and the bacteria. Further work on this tree has shown that many archaebacterial molecules resemble their eukaryotic counterparts more strongly than their bacterial ones and this has led Woese to propose recently[1] that the name 'archaebacteria' should be dropped in favour of 'archae' so as to avoid any possible confusion with the true bacteria.

James Lake uses similar studies based on the same marker

Table 1
What's in a name?
-A Glossary of Terms-

Archae	a 'domain' proposed by Woese comprising all the archaebacteria
Archaebacteria	a group of prokaryotic organisms, each of which lives in extreme environments such as high temperature
Bacteria	a 'domain' proposed by Woese comprising all the eubacteria
Crenarchaeota	sulphur-metabolising, thermoacidophilic archae (Woese)
Domain	a new taxon proposed by Woese which would exist above the level of kingdom
Eocytes	the crenarchaeota (Lake)
Eubacteria	the prokaryotic organisms
Eucarya	a 'domain' proposed by Woese comprising all the eukaryotes
Eukaryotes	the metaphyta, metabionta, protists and fungi - organisms possessing a nucleus and organelles
Euryarchaeota	the halobacteria and methanogens (Woese)
Halobacteria	archaebacteria which live in high ionic-strength environments (e.g. *Halobacterium halobium* which lives in 4M salt)
Holophyletic	phylogenetically grouped with a common ancestor and all of its descendants
Methanogens	archaebacteria which break down carbon compounds to produce methane (e.g. *Methanococcus voltae*)
Monophyletic	holophyletic
Oligophyletic	two or more monophyletic lineages grouped together
Paraphyletic	phylogenetically grouped with a common ancestor and some of its descendants
Taxonomy	the principles of classification

molecule but uses a different algorithm (evolutionary parsimony) an, consequently, produces a different phylogeny - the 'eocyte tree'[2]. In this case, the archaebacteria are not a monophyletic whole but fall into several categories: (i) the eocytes of crenarchaeota which are closely related to the eukaryotes; (ii) the methanogens and (iii) the halophiles which are closely related to the true bacteria.

The third grouping, led by Wolfram Zillig rejects both of the above trees in favour of a fusion hypothesis[3]. In this proposal, the archaebacteria and the bacteria are both paraphyletic lineages and the eukaryotes are regarded as a bi- or oligo-phyletic chimera. Zillig *et al.*'s work is based not upon rRNAs but upon the relationships between eukaryotic, bacterial and archaebacterial DNA-dependant RNA polymerases. Lake believes this work to be invalid[4] because proteins evolve twice as fast as rRNAs but Zillig et al. simply counter that RNA polymerase sequences also contain highly conserved regions and that an alignment of amino acids is inherently more reliable (because less ambiguous) than an alignment of nucleotides.

In a balanced summary of the current state of play in radical thinking, Linkkila and Gogarten[5] acknowledge the importance of using molecular sequences to construct phylogenetic trees but admit that "due to the bias introduced by alignments and algorithms, the obtained significance levels and probabilities for the different trees are not yet sufficient to settle [the argument] unambiguously".

The Conservatives

Standing slightly aloof from this melee are the conservatives, advocates of the retention of either the five-kingdom phylogeny proposed by Whittaker[6] (and later developed by Margulis[8]) or the prokaryote / eukaryote division first noted by Chatton[7] both systems being based upon complex morphologies and large numbers of phenotypic characters. The conservatives believe that, whilst molecular sequence information is a useful classification tool, the radicals over-emphasise its importance at the expense of more traditional, phenotypic methods - a case, perhaps, of not being able to see the phenotypic wood for the molecular trees (sic). Aficionados of the Whittaker system believe that the radicals are wrong to split the prokaryotes into two domains (or whatever) because they all share similar internal organisation and have single genomes and that they are equally wrong when lumping together all eukaryotic organisms (protists, metaphyta, fungi and metazoa) into a single domain. Supporters of the Chatton system (such as Ernst Mayr[9]) agree, but believe that the fundamental division of life is that between prokaryotes and eukaryotes and that this should be the ultimate level of classification below which fall the more traditional, equally ranked, kingdoms (plus the archaebacteria).

The Great Debate

Both the radicals and the conservatives have powerful arguments (and plenty of supporting evidence) for their positions and against the rival view. Radicals believe that theirs is the classification of the future - a brave new world of systematics based on sequences, structures and interrelationships of molecules. They believe that only at the molecular level can one accurately follow the evolutionary process and that phenotypic criteria are now (or soon will be) too limited in scope to be of

much use other than to confirm or enlarge upon conclusions already drawn from molecular sequences. They would deal with the criticism of their splitting of the prokaryotes into two (or more) groups by insisting that the conservatives' defining character of the prokaryotes is essentially a negative one. That is to say, the definition of a prokaryote traditionally is based upon its lacking certain features, e.g. a nucleus. Further, the radicals maintain that if one looks for a positive definition one is forced into their viewpoint: proteins markedly different between the bacteria and the archaebacteria; each possessing unique rRNA structures; and perhaps most tellingly of all, an archaebacterial external architecture which is radically different from either eukaryotes or true bacteria (archaebacterial membranes are composed of ether rather than ester linked isopranyl lipids).

The conservatives would argue that too much emphasis is being based upon too little information. They point out that phylogenetic systems based on a few characters of pre-weighted value (i.e. the rRNAs used by Woese and Lake) have been notoriously misleading. An example of this was the long insistence that the cyanobacteria were eukaryotes (the blue-green algae). Richard Ambler[10] has noted that some evolutionary anomalies in organisms are invariably thrown up by single-gene protein phylogenies and that rRNAs are unlikely to be immune to this.

These anomalies can have various causes such as convergence, parallelism or back-mutation. Sneath[11] further wonders if genes subject to strong selective forces (such as those coding for rRNAs) are in fact, the best examples for studies on phylogeny or whether pseudogenes - which are thought not to experience such pressures as they do not produce functional gene products - might not be better. He asserts that the "logic of such arguments is by no means clear".

Synthesis

As a postgraduate student taking his first tentative steps into this field, it may be a little presumptuous of me to offer an opinion but, of all the hypotheses on offer, that of Zillig's group seems to be the most attractive. In my view it may present a synthesis of the current viewpoints. From the common ancestor there was an original bifurcation point between the archaebacteria and the bacteria. The ancestral eukaryote was then formed as a chimers from the fusion of an archaebacterium and a bacterium. If this was the case, might it not offer an explanation for the curious similarities between the proteins of some eukaryotes to bacterial counterparts and of others to archaebacterial counterparts (particularly Lake's 'eocytes')? Obviously the ideas behind this hypothesis will have to be organised into a coherent whole before trees can be constructed to support them, but the hypothesis is still a neat and satisfying one, especially since an endosymbiotic theory is not a new concept.

As molecular biology chugs onwards, pumping out more and more molecular sequences, a solution to the problem of phylogeny becomes both more complex and more inevitable. To process the volume of data satisfactorily, a pre-requisite is surely an agreement on the best algorithm to use, from to construct phylogenetic trees. Perhaps a sort of "United Nations" of science can be instituted to bring the warring factions together in constructive discussion - if none of the currently used algorithms is satisfactory, co-operation will certainly speed up the development of new and better ones. The final battle, however, will be more semantic than systematic. As Margulis has pointed out[8] even if a phylogenetic tree becomes universally accepted, the classification problem may still prevail, there being several possible defensible positions based on a single tree. I am confident, however, that if we all get together, with a little thought, we can finally decide on a name for the rose. ■

SIMON POTTER
Department of Biochemistry,
University of Edinburgh

References

1. Woese, C. R. *et al.* (1990) *Proc. Natl. Acad. Sci. USA* **87** 4576 - 4579
2. Lake, J. A. (1991) *Trends in Biochem. Sci.* **16** 46 - 50
3. Zillig, W. *et al.* (1989) *Endocytobiosis and Cell Res.* **6** 1 - 25
4. Lake, J. A. (1991) *Trends in Biochem. Sci.* **16** 289 - 290
5. Linkkila, T. P. & Gogarten, J. P. (1991) *Trends in Biochem. Sci..***16** 287 - 288
6. Whittaker, R. H. (1959) *Q. Rev. Biol.* **34** 210 - 226
7. Chatton, E. (1938) *Titres et Travaux Scientifique (1906 - 1937) de Eduoard Chatton*, Sottano Sete
8. Margulis, L. & Guerrero, R. (1991) *New Scientist* **129** 46 - 50
9. Mayr, E. (1990) *Nature* **348** 491
10. Ambler, R. P. (1985) Protein Sequencing and Taxonomy pp.307 - 335 in: M. Goodfellow, D. Jones & F. G. Priest (eds.) *Computer Assisted Bacterial Systematics*, London Academic Press
11. Sneath, P. H. A. (1989) *System Appl. Biol.* **12** 15 - 31

Outrageous! - The Archaebacteria, Phylogeny and Biochemistry

Phylogeny — Recent Ideas

BEFORE the mid-1970s, evolutionary relationships were deduced exclusively from complex morphologies and the fossil record: a process called cladistics. Microbial morphologies, however, are too simple and/or uninterpretable to provide a valid taxonomy, and therefore evolutionary considerations were necessarily confined to complex organisms. Their histories at best cover only 40% of evolutionary time. In the past decade the sequencing revolution has provided a vast store of molecular sequences which can be used to deduce evolutionary relationships between any organisms, whether pro- or eukaryotic. These deductions are made on the basis of studies of molecules known as molecular chronometers or phylogenetic markers. Such a molecule, to be useful, must fulfil the following criteria:

- ubiquity,
- ease of isolation,
- constancy of function (which ensures that any mutations will be selectively neutral).

Molecular chronometers used in the past have indeed fulfilled these criteria. Recently, however, Sneath[1] has questioned their validity for the choice of molecule, and further asks whether phylogeny is better reflected by genes subject to strong selective (i.e. conservative) pressure, or by genes which are subject to low selective pressure. Genes which are constrained by the need to produce a functional product will evolve slowly, while genes which are not so well constrained may be expected to accumulate changes at a steady rate.

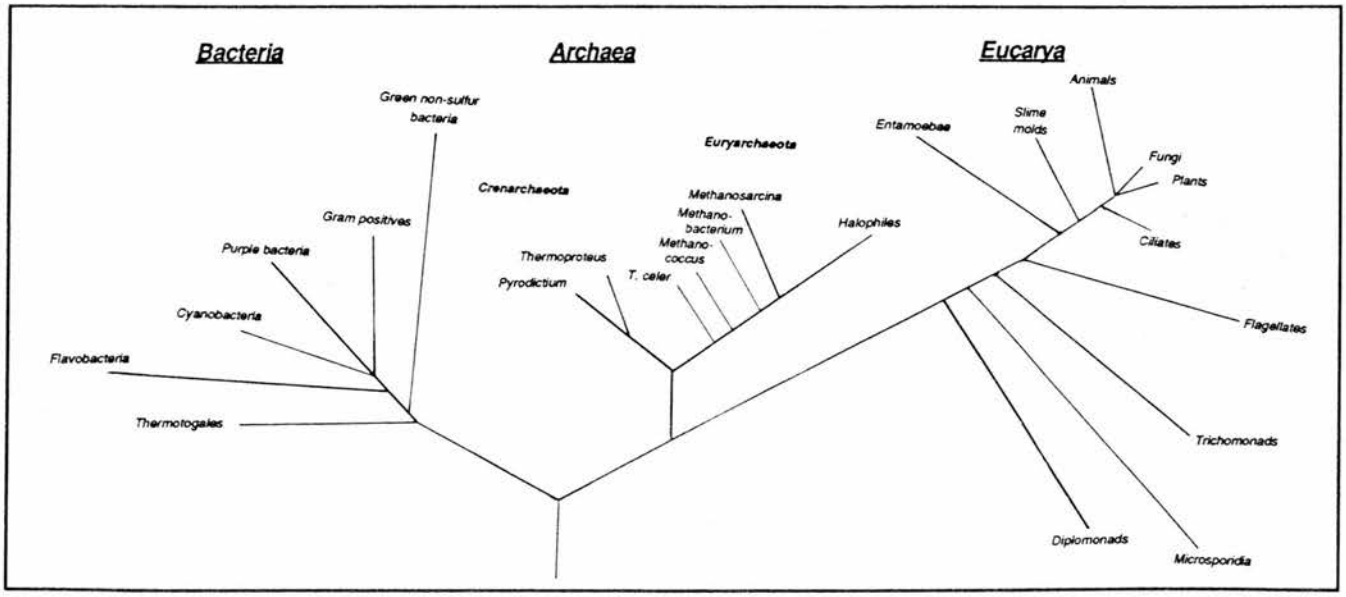
Most of the new phylogeny, however, is based on highly conserved molecules, i.e. proteins and DNA/RNA. Lake[2] has argued that proteins are unsuitable as chronometers because they evolve relatively quickly (particularly those of the eubacteria) and, indeed, most studies up to now have been performed using ribosomal RNA sequences. rRNA sequences are very highly conserved and so even bacterial and archaebacterial sequences can be compared with those of humans. Woese[3] originally used a process known as oligonucleotide cataloguing to construct his trees. 16S/18S rRNA molecules were digested with ribonuclease T1 and the resultant oligonucleotides separated electrophoretically and sequenced. This generated a catalogue of information which could be used to determine the number of nucleotide differences between rRNA sequences. These differences could then be used to construct phylogenetic trees of the type shown in Figure 1. This approach has since been supercveded by the use of complete sequencing studies.

Trees of the type shown in Figure 1 showed up the expected groupings of the eubacteria (or true bacteria) and the eukaryotes (which can be further divided along the lines determined by Whittaker[4]). A surprise, however, was that a third, distinctly different, grouping of organisms was shown to exist by these trees. This grouping comprises all of the archaebacteria, a name which reflects these organisms' proposed antiquity. Archaebacteria are notable for their unique characteristics and for the extreme ecological niches in which they exist. Woese[5] has recently proposed that there should be three domains at the ultimate level of taxonomy: the eucarya (or eukaryotes); the eubacteria (or true bacteria); and the archae (or archaebacteria).

Archaebacteria — A New Domain

Since the 'discovery' of the archaebacteria, much work has been carried out on these organisms, and they have been examined both by biochemical and molecular biological techniques. This work has confirmed the separate status of the archaebacteria, and has shown that while they have many unique properties, in many respects they closely resemble representatives of both the eubacteria and the eucarya. The archaebacteria can be subdivided primarily into three groups: the halophiles, the methanogens, and the sulphur-dependent thermoacidophiles.

Figure 1: Rooted phylogenetic tree constructed from 16S (or 18S) rRNA sequences. Estimates of sequence divergence (mutations fixed per sequence position) were calculated and used by Woese to infer the tree.



Halophiles

These organisms (e.g. *Halobacterium halobium*) have an absolute requirement for high salt concentrations; some even flourish in saturated salt (5.2M) while others have a requirement for high pH (9-10).

Methanogens

These organisms (e.g. *Methanobacterium thermoautotrophicum*) are obligate anaerobes which reduce carbon dioxide to methane. They exist in swampy areas and also occur in some ruminant gut regions.

Sulphur-Dependent Thermoacidophiles

These organisms (e.g. *Sulfolobus solfataricus*) thrive at temperatures as high as 110°C (thought to be the upper limit for organic life), and frequently grow at pH values of 1 or lower. It is worth noting at this point that thermophilicity is the most widespread phenotype throughout the archaebacteria. For example, *Thermoplasma acidophilum* (Figure 2), which grows at 55°C and pH 2, is phylogenetically, if not phenotypically, a methanogen. The phenotype is also found in eubacteria, e.g. *Bacillus stearothermophilus*.

Morphologically, as a group, the archaebacteria are prokaryotic in nature. They are unicellular with no nucleus and have no visible organization within the cell. They do, however, possess some unique structural features. Archaebacterial cell envelopes are completely different from eubacterial ones, lacking muramic acid, diaminopimelic acid, D-glutamic acid, D-alanine and murein. In their place are found *N*-acetylglucosamine/galactosamine and some *L*-amino acids. The predominant glycerolipids in eubacteria and eucarya are glycerides (acylglycerols), whereas in archaebacteria, glycerol ethers are predominant. The presence of C^{40}-tetraethers allows the formation of rigid lipid monolayers instead of the usual bilayers which would melt at high temperatures. Having ether-linked lipids may be a selective advantage because ethers are very resistant to acid/base hydrolysis and may thereby protect the organisms from their hostile environments.

Archaebacteria use the same genetic code as the eubacteria, but some of their genes are known to contain introns. They also contain histone-like proteins, which more closely resemble DNA-binding proteins from eucarya than those from the eubacteria. The initiator tRNAs of methanogens and halophiles resemble those of eubacteria more than those of eucarya, while that of *Sulfolobus* is closer to the eucaryal tRNA; that of *Thermoplasma* is intermediate in nature. Initiation of protein synthesis involves a methionyl-tRNA as in eubacteria.

Detailed study of some of the proteins of the archaebacteria has revealed that they show peculiarities all of their own. This is hardly surprising, as the environments in which they exist would denature 'normal' proteins in a matter of seconds. Some adaptations are therefore to be expected and have indeed been discovered.

Halophilicity

High concentrations of salt in solution denature proteins by withdrawing water from them, causing increased strength of hydrophobic bonds. These strong bonds cause the polypeptide chains of the protein to aggregate and collapse. Halophilic proteins appear to have overcome this problem by having an increased percentage (relative to their mesophilic counterparts) of polar amino acid residues in their structures. These residues compete with the salt for the available water and thereby maintain a hydration shell around the protein — even at high ionic strengths. The levels of the acidic moieties (glutamate and aspartate) are particularly elevated, because they have the highest hydration capacity (6.0-7.5 moles of water / moles of residue) — approximately twice the hydration capacities of lysine and arginine. This high frequency of polar amino acids and consequent low frequency of non-polar residues produces the disadvantage that hydrophobic bonds in the protein are weak and, therefore, the high ambient salt concentration is actually required for the stability of the protein. In the absence of salt, the hydrophobic interactions collapse and the protein unfolds.

Primary structure is not the only adaptation to halophilicity. Special secondary and tertiary structures of the protein may contribute to its binding of water. For example, the malate dehydrogenase from *Halobacterium marismortui*[6] has an active core comprising 80% of the protein which shows a high degree of similarity to the entire eucaryal enzyme (from pig heart). The remaining 20% of the enzyme, not found in the pig heart version, is a large interface with the solvent, and has a highly involuted surface.

Thermophilicity

Thermostable enzymes do not appear to show pronounced differences to their thermolabile counterparts in mesophilic organisms to the disappointment of protein enngineers. From a comparison of the sequences and 3-dimensional structures of haemoglobins from thermophiles and mesophiles. Perutz[7] has suggested that the source of a protein's greater heat stability is an increase in the number of salt bridges on the surface of the protein. It has been shown with the enzyme kanamycin nucleotidyl transferase[8] that one extra salt bridge can significantly increase the heat stability of the protein molecule. There is no associated change in 3-dimensional structure, and only one base change is required to confer the extra thermostability.

As an adjunct to the studies on the proteins and genetic apparatus of the archaebacteria, the pathways of their central metabolic pathways have been studied extensively in the hope that they may shed some light on the evolution of those metabolic routes. For a more comprehensive review than the following summary, see Danson[9].

Glycolysis

The true, generally accepted, central pathway of sugar catabolism is the sequence from triose phosphate to pyruvate. It is found in its entirety in eubacteria, eucarya, the halophiles and, at low level, in some methanogens. Before reaching this common trunk pathway, there is more variation. The Embden-Meyerhof glycolytic pathway (Figure 3) is diagnostic of eucarya and some anaerobic and facultatively anaerobic eubacteria. By this route, glucose is phosphorylated to fructose 1,6-biphosphate, which then undergoes an aldol cleavage to two triose phosphates which enter the trunk pathway. The net energy output of the pathway is 2 x ATP. For a while it was thought that this low-energy output indicated the antiquity of the pathway. A modification of this route, the Entner-Doudoroff pathway, is found in strictly aerobic eubacteria. This change is prompted by the absence in these organisms of 6-phosphofructokinase. The pathway bypasses this enzyme step via 6-phosphogluconate giving an energy output of 1 x ATP, and therefore it is even less efficient than classical glycolysis. The pentose phosphate shunt, which operates alongside these routes in many organisms, is not thought to be a major ATP-generating sequence, but operates to provide the cell with NADPH and pentose and tetrose sugars. Halophiles employ a further modification of the Entner-Doudoroff pathway, in which glucose is oxidized to gluconate and then rejoins the traditional pathway at 2-keto-3-deoxy-6-

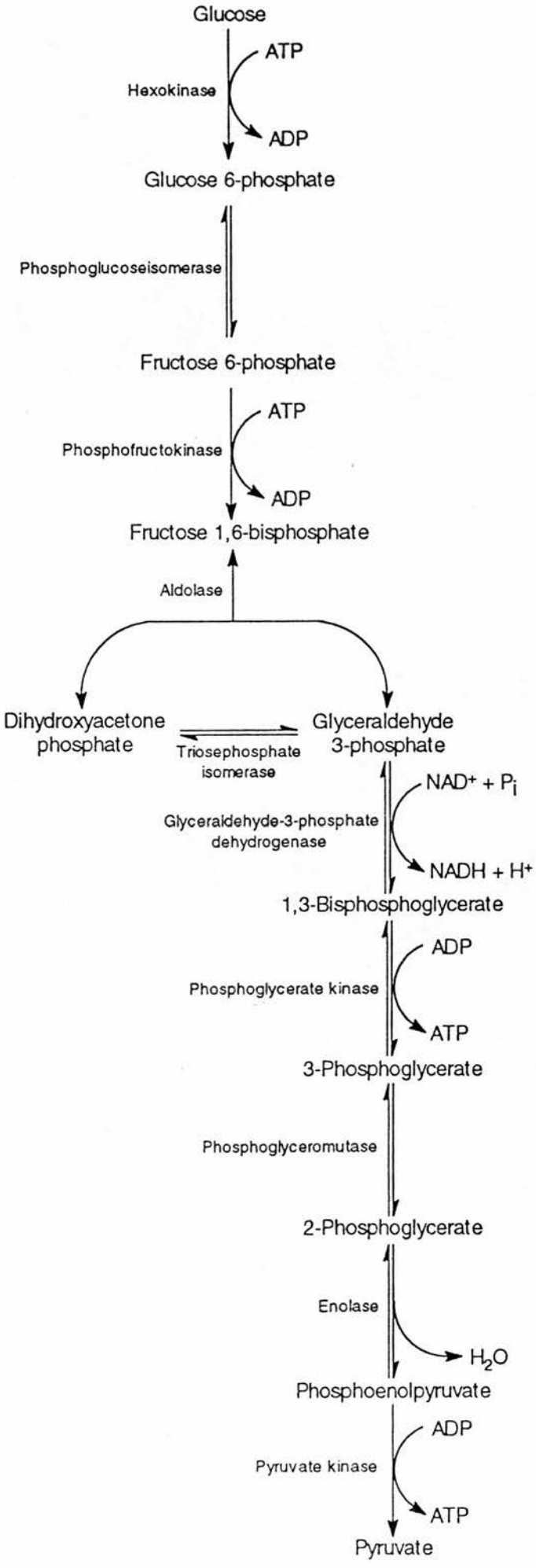


Figure 3. (Left) The Embden-Meyerhof glycolytic pathway

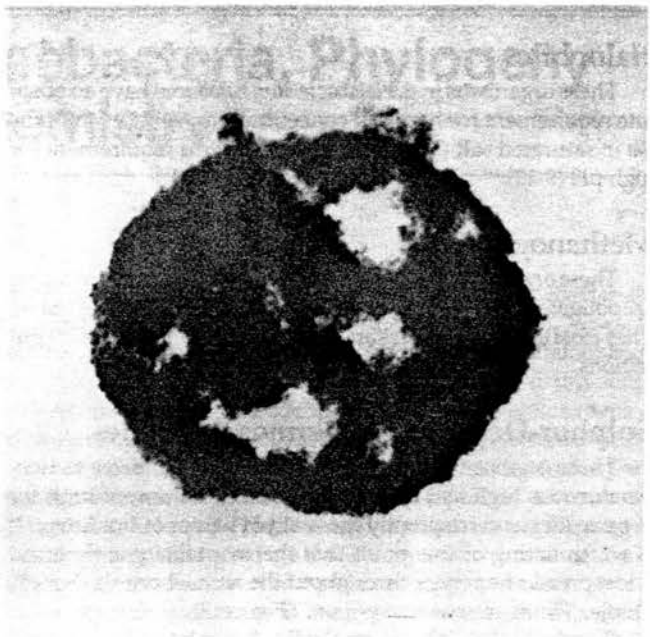


Figure 2. A transmission electron micrograph of Thermoplasma acidophilum.

phosphogluconate. many halophiles, however, use proteins and amino acids rather that carbohydrates as their principle source of carbon. It is the thermophiles which show a more dramatic modification of the classical pathways. Their pathway is non-phosphorylating, and omits almost all of the previously ubiquitous trunk pathway, possessing only two of the traditionally required enzymes — enolase and pyruvate kinase. Radiorespirometric assays indicate that this pathway does occur in *Sulfolobus* and *Thermoplasma* species[10], but there is no net production of ATP. Energy production must therefore occur later in the metabolic pathway. This pathway may suggest that the association of enzymes which constitute the pathway evolved from the 'bottom upwards' and so studies of pyruvate kinase and enolase from these organisms have been considered a priority[11].

Methanogens are autotrophs and produce ATP by the conversion of carbon dioxide and hydrogen into methane. Information about glycolysis is sparse, although it is known that 6-phosphofructokinase is absent and so the Embden-Meyerhof pathway cannot occur. The reverse of the pathway does occur however, converting acetyl-CoA to glucose.

Gluconeogenesis

All archaebacteria possess the capacity for gluconeogenesis from C^3 or C^4 compounds, but the pathways have not yet been established in many organisms, e.g. *Sulfolobus* and *Thermoplasma*. In the halophiles and methanogens, the process is known to be a reversal of the Embden-Meyerhof glycolytic pathway.

Citric Acid Cycle

All archaebacteria can convert pyruvate to acetyl-CoA, but each of the three phenotypes possesses its own distinct version of the subsequent citric acid cycle. The halophiles possess a typical oxidative, aerobic cycle. The situation with the thermophiles, however, is more complex. *T. acidophilum* is an aerobic

Organism	Glucose catabolism	Gluconeogenesis	Citric acid cycle
Halophiles:			
H. saccharovorum	Modified Entner-Doudoroff pathway	Reverse Embden-Meyerhof pathway	Complete oxidative cycle
Thermophiles:			
(1) Sulfolobus	Non-phosphorylated Entner-Doudoroff pathway	Unknown	Autotrophy - reductive cycle; Heterotrophy - oxidative cycle
(2) Thermoplasma	Non-phosphorylated Entner-Doudoroff pathway	Unknown	Complete oxidative cycle
Methanogens:			
Methanobacterium	Entner-Doudoroff pathway	Reverse Embden-Meyerhof pathway	Incomplete reductive cycle

Table 1: Proposed central metabolic pathways of the archaebacteria (after Danson, 1988)

obligate heterotroph and probably possesses an oxidative cycle. *Sulfolobus*, on the other hand, is facultatively autotrophic (i.e. it can grow with carbon dioxide as its sole carbon source) and possesses a reductive cycle. No methanogen has yet been found with a complete citric acid cycle. Two anabolic variations of an incomplete cycle are found instead. *M. thermoautotrophicum* possesses a part cycle (reductive) from oxaloacetate to 2-oxoglutarate. By contrast, *Methanosarcina barkeri* has an incomplete oxidative cycle also leading to 2-oxoglutarate. A summary is shown in Table 1.

As a consequence of these studies, it has been suggested by Danson[9] that the true ancient pathway is the reductive citric acid cycle, perhaps assembled by a primitive progenitor possessing two incomplete arms of the cycle. The advent of oxygen as an electron acceptor then paved the way for the introduction of the oxidative, bioenergetic cycle.

Archaeal futures?

The future importance of studies on the archaebacteria cannot be overemphasized. They provide not only unique insights into the processes of molecular evolution, but they have other, less esoteric uses.

Industrial applications

Halophilicity and thermophilicity are both desirable advantages of a protein molecule in an industrial process. Halophilic proteins could be used at very high substrate concentrations (i.e. low water activity) and thermophilic ones could be used at high reaction temperatures, thus accelerating the process. A recent example of an archaebacterial enzyme which has been commercially marketed is DNA polymerase from *Pyrococcus Woesei*. It is supplied for use in polymerase chain reaction protocols and is thought likely to be much more efficient than the currently used enzyme from *Thermus aquaticus*.

Protein engineering

It may prove feasible to engineer enzymes isolated from mesophilic organisms (based on knowledge gained from archaebacterial proteins) already in commercial use, with a view to increasing their halophilicity and thermostability. Many groups are currently working on projects of this type.

Therapy

The enzyme dihydrolipoamide dehydrogenase has only one known function: as a part of the pyruvate dehydrogenase complex which synthesizes acetyl-CoA. The archaebacterium *H. halobium* does not possess this complex, but has the enzyme occuring free of the complex[12]. Its function there is not yet known. Coincidentally, the enzyme is also present, uncomplexed, in the cell membrane of the parasite *Trypanosoma brucei* which causes sleeping sickness in humans. Studies on the archaebacterial dihydrolipoamide dehydrogenase may help to elucidate a

function for the trypanosome enzyme which might provide an effective therapeutic target.

Implications For Evolution

The type of phylogenetic tree that researchers construct is necessarily dependent on the molecule of choice as a molecular chronometer. Various trees have, therefore, been generated over the past few years, each of which has considerable research evidence as a basis for its construction. Two major trees have been generated, using data from different regions of the same rRNA molecule, from two laboratories which maintain a lively rivalry. The Woese tree (Figure 1) shows the three distinct domains which he proposes, and has been dubbed the archaebacterial tree. Using the same rRNA sequences, Lake proposes a different tree, dubbed the eocyte tree, which groups the sulphur-dependent thermoacidophiles (or eocytes) with the eucarya, and puts the halophiles and methanogens closer to the eubacteria. Lake then renames them photocytes. Each group has refined its particular tree, and neither will accept the other's version - the dispute continues. Both groups agree, however, that only by the study of molecular sequences can a truly accurate phylogeny be constructed.

The above is a 'radical' view. A more 'conservative' view is that held by traditional taxonomists who, while accepting the importance of the evidence provided by molecular sequences, believe that too much emphasis is being placed on it to the detriment of traditional cladistics. The view of these conservatives typified by Ernst Mayr and Lynn Margoulis is that old-style phenotypic evidence cannot be cast aside in favour of the new molecular taxonomy, and that a synthesis of the two bodies of evidence must be achieved. For a more explicit discussion of these viewpoints see Potter[13].

The field is certainly not stagnating, with new trees and their concomitant hypotheses being generated continually. Two recent ones are summarized below

The Zillig 'Fusion' Hypothesis

This hypothesis[14] proposes that the original bifurcation point of the tree of life was between the ancestors of the current eubacteria and the archae. These evolved further until a fusion event occurred between an archaebacterium and a eubacterium. This fusion-generated chimera then went on to evolve into the multifarious species of eucarya. This rather neat idea explains some of the old homologies that the archaebacterial sequences show for both eucaryal and eubacterial sequences, but provides no timescale for the events proposed. The theory is not really a new one, as an endosymbiont hypothesis has already been proposed for the occurrence of chloroplasts and mitochondria in eucaryal cells. Zillig fits these events into his tree as 'secondary invasions'.

The Forterre 'Thermoadaptation' Hypothesis

This hypothesis[15], first expounded at the 640th meeting of the Biochemical Society at Heriot-Watt University in Edinburgh in September 1991, proposes that an ancient organism cannot have been thermophilic because of the RNA world theory of Gilbert[16], namely that RNA is a thermolabile molecule. Hence, Forterre proposes, life arose in mesophilic conditions, and thermophilicity is a relatively recent innovation.

Conclusions

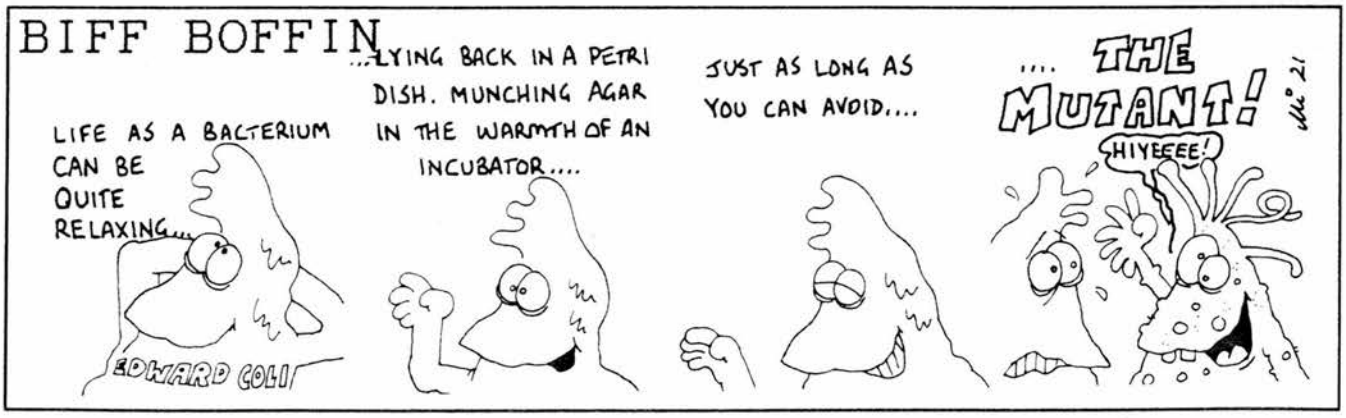
The archaebacteria have great potential: first in helping to understand the processes of evolution before the fossil record begins (with their inclusion, the field of molecular evolution itself is continually evolving, and exciting new developments will doubtless continue for years to come) and secondly, in shaping the future of biotechnology. Even if the archaebacteria could not contribute to such studies, they would still hold a considerable fascination for the biochemist — as Salvador Dali once commented 'the one thing the world will never have enough of is the outrageous'. I think that sums up the archaebacteria very neatly.

For a very comprehensive review of the state of the art in the biochemistry and biotechnology of the archaebacteria, see *The Archaebacteria: Biochemistry and Biotechnology*, (Lunt, G. G., Hough, D. W. & Danson, M. J. eds); Portland Press Ltd., 1992.

SIMON POTTER
University of Edinburgh

References

1. Sneath, P. H. A. (1989) *System. Appl. Biol.* **12**, 15.
2. Lake, J. A. (1991) *Trends Biochem. Sci.* **16**, 46
3. Woese, C. R. & Fox, G. E. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5088
4. Whittaker, R. H. (1959) *Q. Rev. Biol.* **34**, 210
5. Woese, C. R. *et al.* (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576
6. Mevarech, M. & Neumann, E. (1977) *Biochemistry* **16**, 3786
7. Perutz, M. F. (1978) *Science* **201**, 1187
8. Matsumura, M. *et al.* (1984) *J. Bacteriol.* **160**, 413
9. Danson, M. J. (1988) *Adv. Microbial Phys.* **29**, 165
10. Searcy, D. G. & Searcy, J. (1975) *Biochim. Biophys. Acta.* **395**, 535
11. Potter, S. & Fothergill-Gilmore, L. A. (1991) *Biochem. Soc. Trans.* **20**, 11 S
12. Danson, M. J. *et al.* (1986) *Biochemistry* **25**, 3880
13. Potter, S. (1992) *The Biochemist* **14** (1), 21
14. Zillig, W. *et al.* (1989) *Endocytobiosis and Cell Res.* **6**, 1
15. Forterre, P. (1992) in Lunt, G. G., Hough, D. W. & Danson, M. J. (eds.) *The Archaebacteria: Biochemistry and Biotechnology*, Portland Press
16. Gilbert, W. (1986) *Nature* (London) **319**, 618



The Abyss Gazes Also

The Lives and Times of the Early Archaebacteria

Battle not with monsters lest you become a monster.
And remember, as you gaze into the abyss,
the abyss gazes also into you.

Friedrich Nietzsche

Research in the field of molecular evolution is in turmoil. Since I have already outlined the reasons for, and the nature of these difficulties in a previous article in *The Biochemist*[1] suffice it to say here that the most researchers agree on at least one idea. This is that the ancestral cell (or progenote) gave rise initially to two, distinct prokaryotic lineages (and, perhaps simultaneously, the eukaryotic lineage as well). These prokaryotes were the forerunners of the domains we know today as the eubacteria and the archaebacteria. Accepting this, a very interesting question can be posed - did the progenote more closely resemble a eubacterium or an archaebacterium?

To try to answer this question, it is unfortunately necessary to grasp a rather bigger, more stinging nettle - the problem of the origin of life itself. This has become a contentious field of research, with its own journals, peppered with the work of Nobel laureates such as Francis Crick, Thomas Cech and Christian de Duve. Despite the involvement of such luminaries, current thinking in the field is rather muddled - hardly surprising given the timescale of the events being covered and the fact that the first billion years or so of evidence has been eradicated by meteorite impacts. Nevertheless, the paucity of hard evidence hasn't stopped speculation running rampant and three major theories have been proposed.

1. Life from Space

The main exponent of this theory, indeed the only one most people have heard of, is the eclectically inclined astronomer Fred Hoyle[2]. His suggestion - that life was deposited on the young Earth by the meteorite impacts already mentioned - is based on evidence garnered from various sources. Not only have a great number of organic compounds been spectroscopically detected extra-terrestrially (e.g. the tail of Halley's comet is rich in hydrocarbons), but both amino acids and membrane-forming hydrocarbons and alcohols have been found in meteorites called carbonaceous chondrites, which still strike the Earth quite regularly.

Francis Crick, a gentleman somewhat better known for his work down other avenues of research, went a little further than this. He suggested that Arrhenius's 19th century theory of 'panspermia' (basically intact, living organisms which travel through space until they encounter a planetary environment which will support them) was essentially correct, except that they were 'directed' to the Earth by a race of super-intelligent beings[2]. With all due respect, one can only hope that he was joking (indeed, he has since stated as much). A fundamental problem rarely addressed by Hoyle and fellow believers in these concepts is that they provide no clues as to the origins of the extraterrestrial life and hence simply move the problem elsewhere. In summary, these theories, while being quite fun, really detract from the central theme of the current origin of life research - that life arose on the pre-biotic Earth itself.

2. The RNA World

If, like most scientists, you would rather look to the Earth for genesis than into the abyss of space, then matters become more problematic. Ever since 1953 when Stanley Miller become something akin to Mary Shelly's "Modern Prometheus" by generating pre-animate matter (amino acids etc.) from simple ingredients in a flask[3] (see figure 1), scientists have been pondering the problem of how the next step of genesis - from mono-

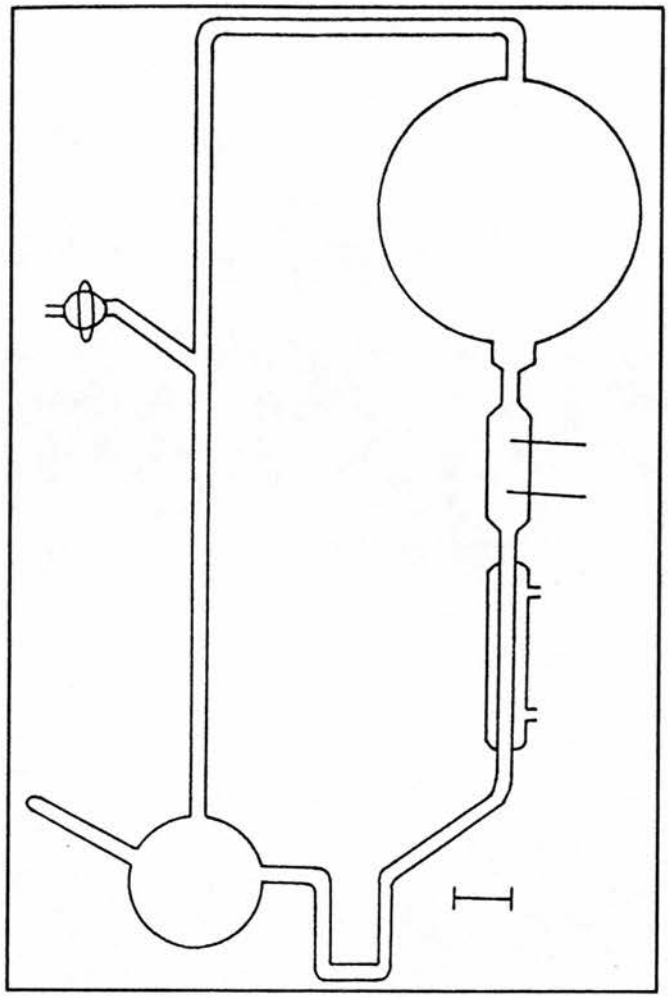


Figure 1: Miller's apparatus. Water was boiled in the smaller flask, mixed with gases in the larger flask, circulated past the electrodes, condensed and emptied back into the boiling flask. The U-tube prevented back circulation. Amino acids accumulated in the water phase. Taken from[3]

mers to self-replacing macromolecules - could have occurred. If you assume that, given enough time, a naturally occurring continuation of Miller's experiment could have produced a molecule of RNA, then a possible scenario presents itself: the ancestral RNA molecules learn how to copy themselves, start to synthesise proteins which serve as catalysts, helping to replicate the RNA and to produce the DNA. The DNA then become the genetic basis for all subsequent life-forms, selfishly creating better 'survival vehicles' for itself and its copies (cf. the theories of Richard Dawkins[4]).

The theory, since christened the 'RNA world hypothesis', was developed by Walter Gilbert of Harvard University after the discovery by Cech and others[5] that certain species of RNA (in *Tetrahymena*) were capable of catalysing self-splicing reactions without the assistance of any protein molecules (see figure 2). They went on to speculate that this may represent a kind of living fossil - a glimpse into a past when RNA performed rather more functions than it does today.

The problem with this theory, which for a while has been enshrined in certain textbooks, is that not only is RNA a rather limited molecule (functionally speaking) in comparison with proteins, but it is notoriously difficult to synthesise *de novo*, particularly under the conditions thought to be prevalent on pre-biotic Earth. Hence, if the RNA world hypothesis is to be believed, life must have arisen in relatively mesophilic conditions (RNA is thermolabile) after a considerable time of cooling of the Earth's surface. Forterre[6], taking this proposal to heart, has suggested that the ancestral progenote was mesophilic in nature and that thermophily (a province extensively populated by the archaebacteria), which was presumed to have been the original phenotype was actually a later innovation.

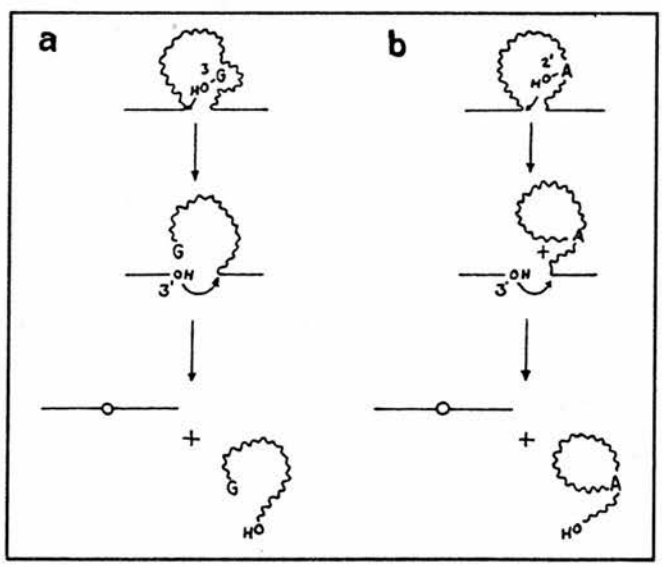
It is, of course, possible that the original replicating molecule was not RNA, although no other convincing candidates have been proposed. In 1990, a self-replacing 'life-like' molecule called the AATE (aminoadenosine triacid ester) was synthesised[7]. When placed in a solution in chloroform with the constituents of the molecule, the AATE can catalyse the synthesis of copies of itself. Interesting though it might be, this experiment has generally been dismissed as irrelevant tinkering, particularly since the replication occurs without making mistakes - a prerequisite for the subsequent evolutionary process.

3. The primary metabolism world

This group of theories (grouped under a title of my own) rejects the central tenet of the RNA world theory - that life could not have developed without a replicating molecule at its heart. Instead, the theories share the common viewpoint that life arose principally as a set of metabolic reactions which went on to produce the replicating molecules that are now the basis of life. In his book "Blueprint for a Cell", Christian de Duve has adapted and enlarged upon the RNA world hypothesis and proposed that thioesters (which arose from the condensation in the pre-biotic soup of carboxylic acids and thiols) acted as the energy source for the formation of multimeric 'proto enzymes' (also formed from thioesters) which then catalysed a cascade of chemical reactions akin to those occurring in modern metabolism[2]. These reactions eventually formed RNA which then precipitated the RNA world.

Other primary metabolism theories have been proposed. A recent one is that of Gunter Wachtershauser[8]. This again states that life evolved as a series of proto-metabolic processes which

Figure 2: Transesterificaton mechanisms for RNA splicing. (a) Self-splicing of Tetrahymena *pre-rRNA and (b) splicing of nuclear pre-mRNA. Intervening sequences are excised as lariats held together by branches containing 2'-5' phosphodiester bonds. Taken from*[5].



occurred on, and were catalysed by, iron pyrite (a metallic mineral containing iron and sulphur: a possible precursor of the FeS complexes found in metabolic processes today). An alternative to iron pyrite as the catalytic surface for proto-metabolism, suggested by A. G. Cairns-Smith of Glasgow University, is a crystalline clay material. This alternative offers the intriguing possibility that the first replicating 'cell' was actually a clay crystal, easily capable of self-replication and evolving in a life-like fashion

For those with an interest in the archaebacteria, these latter theories hold a special attraction. The processes they propose would all very likely have occurred in deep-sea hydrothermal vents (black-smokers), one of the ecological niches occupied almost exclusively by the archaebacteria. Indeed, the de Duve thioester hypothesis depends on an environment of this kind as the synthesis of thioesters requires high temperatures and low pH's. If one accepts this idea then life arose in thermoacidophilic conditions and the first cell would very likely have looked a lot like a modern day archaebacterium.

Abyss gazing

Taking everything discussed above as a guide, it is tempting to try to answer the question originally posed. The following speculation draws together the ideas outlined in this article:

About 4 billion years ago, the Earth is being pummelled by a bombardment of meteorites containing organic compounds of varying complexities. These impacts generate enough heat to boil the surfaces of the oceans and make it highly unlikely that life could arise there. Deep under the ocean surfaces, however, life is in ferment around the hydrothermal vents. There, the progenote replicator, containing a fully integrated DNA-RNA-protein system, has developed via a series of metabolic reactions associated with the surface of iron pyrite crystals in the presence of thioester catalysts. The replicators (initially pro-archaebacterial in nature) make enough mistakes to allow evolution to take place and two major types of prokaryotic organisms form [Note: Prokaryotes would have been more likely to survive because the half-life of RNA would have been quite short under the conditions prevalent around the vent and hence it would be a useful molecule only in small, compact organisms]. Both forms spread to cover the Earth giving us both the eubacteria and archaebacteria that we are familiar with today. Much later, a genomic catastrophe occurs fusing the genomes of a eubacterium and an archaebacterium to give the progenote eukaryote. The three domains of life have now formed.

If one accepts this scheme, which admittedly suffers from a lot of drawbacks (e.g. black-smokers tend to have short lifetimes and, therefore, time might be a problem; large portions of the scheme are based on pure speculation; the exact effects of extensive meteorite bombardments are difficult to determine etc.), then it does seem likely that the progenote was archaebacterial in phenotypic terms (and, therefore, presumably phylogenetically too). It does offer a plausible set of proposals for the genesis of life, most stages of which are testable in the laboratory.

Whatever one thinks of this particular scheme, it is possible that something quite like it will eventually become a central dogma of biological sciences, just as the RNA world hypothesis latterly became. Comment is invited! Furthermore, it is worth remembering that, whatever the answer is to the riddle of the origin of life on Earth, it is a part of us. We are made up of essentially the same stuff as the first replicators 4 billion years ago. Therefore, as we gaze into the abyss of evolutionary time, we must bear in mind that the abyss is, in fact, hidden with ourselves.

Finally, recent research has shown that RNA catalysis is not quite as limited in scope as previously thought. This lends support to the RNA World hypothesis[9].

SIMON POTTER
Department of Biochemistry
University of Edinburgh

References:

1. Potter, S. (1992) *The Biochemist* **14**(1) 21-24
2. de Duve, C. (1991) *Blueprint for a Cell* Neil Patterson Publishers
3. Miller, S. L. (1953) *Science* **117** 528-529
4. Dawkins, R. (1989) *The Selfish Gene* (New Edition) Oxford University Press
5. Cech, T. R. and Bass, B. (1986) *Ann. Rev. Biochem.* **55** 599-629
6. Forterre, P. (1992) in: Lunt, G. G., Hough, D. W. and Danson, M. J. (eds.) *The Archaebacteria: Biochemistry and Biotechnology* Portland Press Ltd
7. Tjivikua, T. *et al.* (1990) *J. Am. Chem. S.* **112** 1249-1250
8. Wachtershauser, G. (1990) *Proc. Natl. Acad. Sci* USA **87** 200-204
9. Noller, H. F. *et al.* (1992) *Science*, **256**, 1416-1419.